JMIR Formative Research

Impact Factor (2023): 2.0
Volume 6 (2022), Issue 6 ISSN 2561-326X Editor in Chief: Amaryllis Mavragani, PhDc

Contents

Original Papers

Short Paper

Viewpoint

Letters to the Editor

Original Paper

Continuation of Teletherapy After the COVID-19 Pandemic: Survey Study of Licensed Mental Health Professionals

Rashmi Gangamma[1], PhD; Bhavneet Walia[2], PhD; Melissa Luke[3], PhD; Claudine Lucena[1], MA

[1]Department of Marriage and Family Therapy, Falk College of Sport and Human Dynamics, Syracuse University, Syracuse, NY, United States

[2]Department of Public Health, Falk College of Sport and Human Dynamics, Syracuse University, Syracuse, NY, United States

[3]Department of Counseling & Human Services, School of Education, Syracuse University, Syracuse, NY, United States

Corresponding Author:
Rashmi Gangamma, PhD
Department of Marriage and Family Therapy
Falk College of Sport and Human Dynamics
Syracuse University
Peck Hall
601 E Genesee Street
Syracuse, NY, 13202
United States
Phone: 1 614 325 4029
Email: rgangamm@syr.edu

Abstract

Background: The use of teletherapy has exponentially increased in the context of the ongoing COVID-19 pandemic. Studies on teletherapy documented substantial benefits of accessibility and convenience even before the start of the pandemic. Although recent studies show that this modality of therapy delivery is here to stay, few have studied who will most benefit from this trend.

Objective: In this paper, we report predictors of continued teletherapy usage in a sample of licensed mental health professionals in the United States during a time period when pandemic-related restrictions began diminishing. As such, it is one of the first studies to examine factors related to continued benefits of teletherapy postpandemic.

Methods: Participation from licensed mental health professionals was sought on listservs of national organizations of multiple mental health organizations. Data were collected via an anonymous link to a survey on Qualtrics between January 2021 and April 2021. Participants responded to questions on therapist demographics, practice setting, experiences of shifting to teletherapy, perspectives on continued use of teletherapy, and their client characteristics. Findings related to client characteristics that predicted continued teletherapy usage are presented here.

Results: A total of 186 individuals consented to participate in the survey, with a final sample of 114 with complete data. A majority of participants identified as female (92/114, 80.7%), White (94/114, 82.5%), and having a master's degree (75/114, 65.5%) from a nationally accredited program (106/114, 93%). Data were analyzed using heteroskedastic regression modeling with client-related factors as predictors. Two models were run with and without distance travelled by clients as a control variable. Model estimates from both models showed that continued use of teletherapy postpandemic was predicted by the following factors: higher percentage of clients from rural areas, younger and older adult clients, clients with Medicare, and clients with marginalized gender and religious/spiritual identities. Significantly, having a higher percentage of clients from lower socioeconomic status, a higher percentage of those with Medicaid coverage, and a higher percentage of couples and families as clients predicted decreased use of teletherapy postpandemic.

Conclusions: Findings from the study suggest that while some groups of clients are more likely to continue to receive benefits of teletherapy, vulnerable groups such as those in lower socioeconomic conditions, Medicaid beneficiaries, and those who seek couple and family therapy may be less likely to be served by it. These differences point to a need to address factors driving telehealth care disparities such as access to technology, housing, and childcare issues, as well as the need for continued training for licensed professionals.

(JMIR Form Res 2022;6(6):e32419) doi:10.2196/32419

KEYWORDS

teletherapy; relational teletherapy; teletherapy predictors; postpandemic teletherapy; mental health; telemedicine; COVID-19; telehealth

Introduction

The COVID-19 pandemic and subsequent social measures drastically impacted society [1], shifting education, work, health care [2,3], and mental health [4]. Telemental health, referred to as teletherapy, has been used over the past 20 years [5] with demonstrated effectiveness [6,7]. Teletherapy refers to the use of electronically based communication such as videoconferencing, telephone calls, and mobile apps to provide access to mental health services, typically across distances [8]. Rapid legislative changes, training, and guidelines resulted in an exponential increase in teletherapy when compared to prepandemic levels [9,10]. The increase in relational teletherapy (teletherapy with couples and families) has been particularly important given increased risks for distress, anxiety, grief/loss, substance abuse, and family violence in children [11] and adults [12-14] during the pandemic. Before the COVID-19 pandemic, scholars contended that historically underserved populations derived more benefits from the flexibility and accessibility of teletherapy [15,16]. As COVID-19–related restrictions are lifted, teletherapy will remain part of the mental health landscape [17]. However, given the existing challenges of the need for training, technological advances, and other barriers to effective use [8,18,19], we are yet to understand whether teletherapy will be accessible equitably postpandemic.

In this paper, we present findings from a study on predictors of continued teletherapy practice postpandemic from a sample of licensed mental health practitioners. Specifically, our research question was “What factors of therapist practice predict their intention for continued use of teletherapy practice postpandemic?” Existing literature suggests that distance from services, client profile [15,16], and vulnerability of selected client populations [6-8,18-20] may influence provision of teletherapy. Clarifying predictors would strengthen recent research on therapists’ experiences transitioning to the use of telehealth [18] and may assist in identifying factors in disparities in telehealth care postpandemic.

Methods

Recruitment

Participation was open to licensed mental health professionals who were currently providing teletherapy. Upon institutional review board approval, a link to an anonymous Qualtrics survey was posted on multiple listservs including the American Association for Marriage and Family Therapy, the American Counseling Association, as well as professional groups for social workers. Data were gathered between January 2021 and April 2021, when increased vaccinations were driving gradual removal of public health reductions [20]. Survey questions included therapist demographics, practice setting, experiences of shifting to teletherapy, perspectives on continued teletherapy use, and client characteristics. No incentives were provided; instead, a donation was made to a nonprofit chosen by participants. A total of 186 individuals consented to participate in the survey, with a final sample of 114 with complete data.

Ethics Approval

This study received ethics approval from Syracuse University’s Institutional Review Board (IRB #20-310).

Statistical Analysis

Descriptive statistics and regression analyses were conducted using Stata software (version 14; StataCorp LLC) [21]. A residual plot revealed increasing standard deviation of residuals in the independent variables (ie, heteroskedasticity). Given that errors were normally distributed and mean and variance functions were correctly specified, we ran hetregress regression models with maximum likelihood estimator [21]. Using G*Power power analysis, setting a medium effect size with 10 predictors in our model, we determined that our final sample of 114 was sufficient for regression analysis [22].

Results

Participants were from 27 states in the United States, with a majority identifying as female (92/114, 80.7%), White (94/114, 82.5%), and with a master's degree (75/114, 65.5%) from a nationally accredited program (106/114, 93%). Less than half of participants (45/114, 39.5%) reported prepandemic experience practicing teletherapy. Table 1 shows other practice profiles of participants and Table 2 shows client profile factors used as independent variables in the regression models.

Table 3 shows coefficient values of regression models run without and with control for distance travelled by clients (models 1 and 2, respectively). We controlled for distance from a health setting in model 2 to limit multicollinearity and increase robustness of estimates. Both models were estimated with therapist gender as a cluster variable.

Among factors examined, statistically significant predictors were (1) higher percentage of clients living further from a metro area, particularly those in rural areas (β=38.578, P<.01), (2) higher percentage of clients who are younger (<30 years; β=.186, P<.001) or older (65-80 years; β=.634, P<.001), (3) higher percentage of clients who identified with a minoritized gender (β=.223, P<.001) and religious/spiritual identity (β=.153, P<.001), and those with disabilities (β=.399, P<.001), and (4) higher percentage of clients with Medicare (β=.457, P<.001).

Conversely, therapists for whom couples/families were >75% of their caseload were less likely to continue teletherapy compared to therapists with caseloads of couples/families <25% (β=19.876, P<.001), 25%-50% (β=32.040, P<.001) and 50%-75% (β=28.927, P<.001). Similarly, therapists with a higher percentage of clients from lower socioeconomic backgrounds (β=–.285, P<.001) and a higher percentage of clients with Medicaid coverage (β=–.143, P<.05) were less likely to continue teletherapy postpandemic.

Table 1. Practice profiles of participants (N=114).

Practice profile of participants	Participants, n (%)
Type of license	
Marriage and family therapy	77 (67.5)
Mental health counselor	21 (18.2)
Clinical social work	5 (4.4)
Clinical psychologist	4 (3.5)
Other	7 (6.1)
Geographical location	
Large metro	36 (31.9)
Medium metro	32 (28.3)
Small metro	27 (23.9)
Rural area	6 (5.3)
Small town	5 (4.4)
Distance travelled by clients	
<25 miles	98 (85.8)
25-50 miles	13 (11.5)
>50 miles	3 (2.4)

Table 2. Descriptive of client profile factors used in regression models.

Client profile	Average percentage[a]
Age group (years)	
<30	44.05
30-49	38.75
50-64	10.83
65-80	4.20
>80	0.34
Gender	
Female	56.42
Male	34.81
Nonbinary/gender expansive	5.19
Transgender	4.81
Other	1.39
Marginalized identities	
Marginalized gender identities	15.22
Marginalized sexual identities	17.79
Marginalized racial/ethnic identities	26.22
Marginalized religious/spiritual identities	10.01
Lower socioeconomic status groups	28.38
Having a disability	15.91
Veterans	5.96
Payer mix	
Medicaid	13.01
Medicare	4.42
Private health insurance	27.81
Veterans Health Care	2.19
Self-pay	43.71
Other	8.63
Percentage of couples and families in case load	
<25%	42.98
25%-50%	0.34
50%-75%	11.40
>75%	12.28

[a]Absolute values are unavailable because the average percentage was calculated for each group.

Table 3. Regression model of client factors predicting therapists' postpandemic teletherapy usage.

Factors	Model 1 (n=94)		Model 2 (n=94)	
	Coefficient	SE	Coefficient	SE
Practice setting				
Fringe large metro	6.792	0.436	9.670	0.499
Medium metro	7.495[a]	3.418	5.545[b]	1.876
Small metro	6.620[a]	3.960	5.401	0.928
Micropolitan	16.804[a]	2.804	15.939[a]	3.028
Rural	39.843[c]	1.970	38.578[c]	2.079
Percentage of couples and families in case load				
<25%	25.291[a]	3.518	19.876[a]	2.993
25%-50%	39.158[a]	29.207	32.040[a]	9.333
50%-75%	35.416[a]	5.746	28.927[a]	4.351
Client age (years)				
<30	0.213[a]	16.047	0.186[a]	7.052
30-49	0.277[a]	28.157	0.226[a]	5.083
51-64	–0.215	–0.655	–0.135	–0.365
65-80	0.661[c]	2.468	0.634[a]	2.961
Percentage of clients with marginalized identities				
Racial/ethnic identities	0.089	0.921	0.134	1.129
Sexual identities	0.005	0.033	0.009	0.079
Gender identities	0.276[a]	4.766	0.223[a]	6.154
Religious/spiritual identities	0.109[c]	2.069	0.153[b]	1.855
Lower socioeconomic status	–0.341[a]	–3.879	–0.285[a]	–3.264
Disability	0.417[a]	6.261	0.399[a]	3.734
Client payment modality				
Medicaid	–0.066	–0.871	–0.143[b]	–1.649
Medicare	0.390[a]	4.139	0.457[a]	4.823
Private insurance	–0.071[a]	–4.344	–0.079[a]	–3.712
Other pay	0.148[a]	3.151	0.090[a]	2.787
Constant	–83.033[a]	–6.727	–87.333[a]	–6.786
Insigma 2 Constant	6.068[a]	58.085	6.161[a]	54.540

[a]$P<.001$.
[b]$P<.05$.
[c]$P<.01$.

Discussion

Principal Findings

Results illuminate the potential types of clients most likely to continue to receive teletherapy postpandemic from licensed mental health professionals in our sample. In addition to supporting earlier literature on use of teletherapy with clients with disabilities and from rural areas [23,24], our findings suggest that younger and older adult clients, those on Medicare, and clients who identified with marginalized gender or religious/spiritual identities are most likely to continue to receive teletherapy. It is likely that legislative actions leading to waivers of restrictions and increased coverage of teletherapy [25,26]

benefitted older adult clients and those with Medicare coverage. For clients with minoritized social identities who could also access teletherapy, changes during the pandemic may have highlighted the relative safety of seeking therapy via technology.

We also found that therapists were less likely to continue teletherapy when they had a higher percentage of clients from lower socioeconomic backgrounds and with Medicaid coverage or had a higher percentage of caseloads with couples and families. Given that the pandemic has disproportionately impacted those who are underresourced, decreased teletherapy usage with those with lower socioeconomic status suggests that unless structural issues of accessibility are addressed, vulnerable groups may be left behind. Studies report technological difficulties, lack of confidential space, and privacy concerns hinder relational teletherapy [27]. It is possible these barriers are indicative of a need for structural changes (eg, access to adequate housing, broadband internet, and childcare) to prevent deepening disparities. Although therapists with a higher percentage of Medicare clients were likely to continue its use, those with a greater percentage of Medicaid clients were less likely to do so. Given both Medicare and Medicaid coverage of teletherapy began at the same time, this difference may be a factor of available client resources or discrepancies in support between the two programs at state and local levels.

Another significant finding is therapists with the highest percentage of couples and families in their caseload were less likely to continue teletherapy. Although we did not ask for their reasons, this is consistent with earlier studies identifying challenges of training [8], difficulties in de-escalating, and simultaneous engagement with multiple family members [28]. Although teletherapy presents several advantages for access with partners in multiple locations or families with young children [7,18,27], COVID-19 factors related to remote work and school, limited space at home, and lack of social support may have resulted in intense situations [29] that were challenging to address via teletherapy. Studies have reiterated these challenges, including the possibility of therapist exhaustion [30], moral distress [31], split alliances [18], and lack of training and competencies in teletherapy [8]. Moving forward, competency-based training [19] and best practices for telemental health must attend to the unique challenges of working with couples and families [27] along with ways in which therapists can be better supported [32]. Further research is also needed to better differentiate therapists' experiences with telehealth in general from their unique experiences of teletherapy during the COVID-19 pandemic [18].

Limitations

Although this study recruited from different states and mental health disciplines, and the findings are robust, they are still exploratory and tentative. Participants self-selected to take part in the survey, and it is possible they had specific experiences that may not reflect views of the national population of therapists, limiting generalizability. Future research with a diverse sample and increased heterogeneity is needed. Doing so may result in less heteroscedastic data and extend our understanding of how aspects of the therapist, client, and practice contexts intersect.

Conclusion

Public health concerns and health safety underscored the shift to teletherapy [33], rather than a structured or clinically sound plan to increase access with trained practitioners. As we emerge from pandemic-related restrictions, it is likely that teletherapy will continue [17]. However, few studies have examined mental health providers' perspective on potential inequities of shifting to teletherapy [34] and the resultant disproportionate experiences of those living in underresourced communities [35]. Although access and convenience drive teletherapy use [36], our study suggests that after the pandemic, licensed professionals are less likely to continue teletherapy for clients in lower socioeconomic groups as well as for many couples and families. We contend that training clinicians and addressing structural barriers to teletherapy access may decrease deepening disparities in teletherapy provision.

Conflicts of Interest

None declared.

References

1. Van Lancker W, Parolin Z. COVID-19, school closures, and child poverty: a social crisis in the making. Lancet Public Health 2020 May;5(5):e243-e244 [FREE Full text] [doi: 10.1016/S2468-2667(20)30084-0] [Medline: 32275858]
2. Holmes E, O'Connor RC, Perry V, Tracey I, Wessely S, Arseneault L, et al. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. Lancet Psychiatry 2020 Jun;7(6):547-560 [FREE Full text] [doi: 10.1016/S2215-0366(20)30168-1] [Medline: 32304649]
3. Walia B, Shridhar A, Arasu P, Singh GK. US Physicians' Perspective on the Sudden Shift to Telehealth: Survey Study. JMIR Hum Factors 2021 Aug 12;8(3):e26336 [FREE Full text] [doi: 10.2196/26336] [Medline: 33938813]
4. Luiggi-Hernández J, Rivera-Amador A. Reconceptualizing Social Distancing: Teletherapy and Social Inequality During the COVID-19 and Loneliness Pandemics. Journal of Humanistic Psychology 2020 Jul 03;60(5):626-638 [FREE Full text] [doi: 10.1177/0022167820937503]
5. TeleMental Health in the Department of Veteran Affairs. U.S. Department of Veterans Affairs. URL: https://www.va.gov/anywheretoanywhere/docs/TeleMental_Health_factsheet.PDF [accessed 2021-07-20]
6. Rees C, Maclaine E. A Systematic Review of Videoconference - Delivered Psychological Treatment for Anxiety Disorders. Australian Psychologist 2020 Nov 12;50(4):259-264 [FREE Full text] [doi: 10.1111/ap.12122]

7. Wrape E, McGinn M. Clinical and Ethical Considerations for Delivering Couple and Family Therapy via Telehealth. J Marital Fam Ther 2019 Apr;45(2):296-308 [FREE Full text] [doi: 10.1111/jmft.12319] [Medline: 29361194]
8. Pickens J, Morris N, Johnson D. The Digital Divide: Couple and Family Therapy Programs' Integration of Teletherapy Training and Education. J Marital Fam Ther 2020 Apr;46(2):186-200. [doi: 10.1111/jmft.12417] [Medline: 31820834]
9. McKee G, Pierce B, Tyler C, Perrin P, Elliott T. The COVID-19 Pandemic's Influence on Family Systems Therapists' Provision of Teletherapy. Fam Process 2022 Mar;61(1):155-166 [FREE Full text] [doi: 10.1111/famp.12665] [Medline: 33913526]
10. Mishna F, Milne E, Bogo M, Pereira L. Responding to COVID-19: New Trends in Social Workers' Use of Information and Communication Technology. Clin Soc Work J 2021;49(4):484-494 [FREE Full text] [doi: 10.1007/s10615-020-00780-x] [Medline: 33250542]
11. Golberstein E, Wen H, Miller B. Coronavirus Disease 2019 (COVID-19) and Mental Health for Children and Adolescents. JAMA Pediatr 2020 Sep 01;174(9):819-820 [FREE Full text] [doi: 10.1001/jamapediatrics.2020.1456] [Medline: 32286618]
12. Cluver L, Lachman J, Sherr L, Wessels I, Krug E, Rakotomalala S, et al. Parenting in a time of COVID-19. Lancet 2020 Apr 11;395(10231):e64 [FREE Full text] [doi: 10.1016/S0140-6736(20)30736-4] [Medline: 32220657]
13. Hammonds C, Kerrissey J, Tomaskovic-Devey D. Stressed, Unsafe, and Insecurential Workers Need A New, New Deal. University of Massachusetts Amherst Labor Center and Center for Employment Equity. 2020. URL: https://www.umass.edu/employmentequity/stressed-unsafe-and-insecure-essential-workers-need-new-new-deal [accessed 2021-07-20]
14. Trzebiński J, Cabański M, Czarnecka J. Reaction to the COVID-19 Pandemic: The Influence of Meaning in Life, Life Satisfaction, and Assumptions on World Orderliness and Positivity. Journal of Loss and Trauma 2020 May 21;25(6-7):544-557 [FREE Full text] [doi: 10.1080/15325024.2020.1765098]
15. Goss C, Richardson W, Dailey N, Bair B, Nagamoto H, Manson SM, et al. Rural American Indian and Alaska Native veterans' telemental health: A model of culturally centered care. Psychol Serv 2017 Aug;14(3):270-278 [FREE Full text] [doi: 10.1037/ser0000149] [Medline: 28805411]
16. Snow W, Lamar M, Hinkle J, Speciale M. Current Practices in Online Counselor Education. TPC 2018 Jun;8(2):131-145 [FREE Full text] [doi: 10.15241/whs.8.2.131]
17. Sheperis D, Smith A. Telehealth Best Practice: A Call for Standards of Care. JTCES 2021;1(1):27-35 [FREE Full text] [doi: 10.22371/tces/0004]
18. Orlowski E, Friedlander M, Megale A, Peterson EK, Anderson SR. Couple and family therapists' experiences with Telehealth during the COVID-19 pandemic: a phenomenological analysis. Contemp Fam Ther 2022;44(2):101-114 [FREE Full text] [doi: 10.1007/s10591-022-09640-x] [Medline: 35400810]
19. Springer PR, Bischoff RJ, Taylor NC, Neuhaus V, Leow C. Competency-based training in the supervision of relational telemental supervision. J Marital Fam Ther 2021 Apr 29;47(2):375-391. [doi: 10.1111/jmft.12513] [Medline: 33780568]
20. Schuchat A, Walensky R. Summary of Guidance Review. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/downloads/communication/Guidance-Review.pdf [accessed 2022-05-25]
21. Stata 14. StataCorp. 2015. URL: https://www.stata.com/ [accessed 2022-05-25]
22. G*Power. URL: http://www.gpower.hhu.de/ [accessed 2022-04-14]
23. Hines M, Bulkeley K, Dudley S, Cameron S, Lincoln M. Delivering Quality Allied Health Services to Children with Complex Disability via Telepractice: Lessons Learned from Four Case Studies. J Dev Phys Disabil 2019 Feb 16;31(5):593-609 [FREE Full text] [doi: 10.1007/s10882-019-09662-8]
24. Turgoose D, Ashwick R, Murphy D. Systematic review of lessons learned from delivering tele-therapy to veterans with post-traumatic stress disorder. J Telemed Telecare 2018 Oct;24(9):575-585. [doi: 10.1177/1357633X17730443] [Medline: 28958211]
25. Rogers J. Coding telehealth services during COVID-19. Nurse Pract 2021 Feb 01;46(2):10-12 [FREE Full text] [doi: 10.1097/01.NPR.0000731584.40074.eb] [Medline: 33399335]
26. Lowey NM. Text - H.R.6074 - 116th Congress (2019-2020): Coronavirus Preparedness and Response Supplemental Appropriations Act, 2020. US Congress. URL: https://www.congress.gov/bill/116th-congress/house-bill/6074/text [accessed 2021-07-20]
27. Hardy N, Maier C, Gregson T. Couple teletherapy in the era of COVID-19: Experiences and recommendations. J Marital Fam Ther 2021 Apr;47(2):225-243 [FREE Full text] [doi: 10.1111/jmft.12501] [Medline: 33742712]
28. McCoy M, Hjelmstad L, Stinson M. The Role of Tele-Mental Health in Therapy for Couples in Long-Distance Relationships. Journal of Couple & Relationship Therapy 2013 Oct;12(4):339-358 [FREE Full text] [doi: 10.1080/15332691.2013.836053]
29. Lebow J. Family in the Age of COVID-19. Fam Process 2020 Jun;59(2):309-312 [FREE Full text] [doi: 10.1111/famp.12543] [Medline: 32412686]
30. Heiden-Rootes K, Ferber M, Meyer D, Zubatsky M, Wittenborn A. Relational teletherapy experiences of couple and family therapy trainees: "Reading the room," exhaustion, and the comforts of home. J Marital Fam Ther 2021 Apr;47(2):342-358 [FREE Full text] [doi: 10.1111/jmft.12486] [Medline: 33449401]
31. Patterson JE, Edwards TM, Griffith JL, Wright S. Moral distress of medical family therapists and their physician colleagues during the transition to COVID-19. J Marital Fam Ther 2021 Apr 09;47(2):289-303 [FREE Full text] [doi: 10.1111/jmft.12504] [Medline: 33751596]

32. Hertlein K, Drude K, Jordan S. "What Next?": Toward telebehavioral health sustainability in couple and family therapy. J Marital Fam Ther 2021 Jul;47(3):551-565 [FREE Full text] [doi: 10.1111/jmft.12510] [Medline: 33818791]
33. Eppler C. Systemic teletherapists' meaningful experiences during the first months of the coronavirus pandemic. J Marital Fam Ther 2021 Apr;47(2):244-258 [FREE Full text] [doi: 10.1111/jmft.12515] [Medline: 33774847]
34. Velasquez D, Mehrotra A. Ensuring The Growth Of Telehealth During COVID-19 Does Not Exacerbate Disparities In Care. Health Affairs Blog. URL: https://www.healthaffairs.org/do/10.1377/hblog20200505.591306/full/ [accessed 2022-05-25]
35. Devakumar D, Shannon G, Bhopal S, Abubakar I. Racism and discrimination in COVID-19 responses. Lancet 2020 Apr 11;395(10231):1194 [FREE Full text] [doi: 10.1016/S0140-6736(20)30792-3] [Medline: 32246915]
36. Stoll J, Müller JA, Trachsel M. Ethical Issues in Online Psychotherapy: A Narrative Review. Front Psychiatry 2019;10:993 [FREE Full text] [doi: 10.3389/fpsyt.2019.00993] [Medline: 32116819]

Edited by A Mavragani; submitted 27.07.21; peer-reviewed by K Hertlein, S Gordon, D Sheparis; comments to author 01.03.22; revised version received 25.04.22; accepted 02.05.22; published 01.06.22.

Please cite as:
Gangamma R, Walia B, Luke M, Lucena C
Continuation of Teletherapy After the COVID-19 Pandemic: Survey Study of Licensed Mental Health Professionals
JMIR Form Res 2022;6(6):e32419
URL: https://formative.jmir.org/2022/6/e32419
doi:10.2196/32419
PMID:35584317

Original Paper

Direct Outreach in Bars and Clubs to Enroll Cigarette Smokers in Mobile Cessation Services: Exploratory Study

Patricia Chalela[1*], DrPH; Alfred L McAlister[2*], PhD; Cliff Despres[1*], BJ; Edgar Muñoz[1*], MS; Pramod Sukumaran[1*], PhD; David Akopian[3*], PhD; Sahak Kaghyan[3*], PhD; Jesus Trujillo[4*], BA; Amelie G Ramirez[1*], DrPH

[1]Institute for Health Promotion Research, University of Texas Health Science Center at San Antonio, San Antonio, TX, United States

[2]School of Public Health, Austin Regional Campus, University of Texas Health Science Center at Houston, Austin, TX, United States

[3]College of Engineering, The University of Texas at San Antonio, San Antonio, TX, United States

[4]Foundry512, Austin, TX, United States

[*]all authors contributed equally

Corresponding Author:
Amelie G Ramirez, DrPH
Institute for Health Promotion Research
University of Texas Health Science Center at San Antonio
Suite 1000
7411 John Smith Drive
San Antonio, TX, 78229
United States
Phone: 1 210 562 6500
Fax: 1 210 562 6545
Email: ramirezag@uthscsa.edu

Abstract

Background: Cigarette smoking and alcohol use are well known to be concomitant behaviors, but there is a lack of studies related to recruitment of smokers for mobile cessation services at places where alcohol is consumed, such as bars and clubs. Adapting recruitment strategies to expand the reach of cessation programs to where tobacco users are located may help decrease the health-equity gap in tobacco control by improving reach and enrollment of underserved smokers residing in low-income and rural areas who are not reached by traditional cessation services.

Objective: The purpose of this exploratory study was to assess the feasibility of direct outreach in bars, clubs, and restaurants to recruit smokers to Quitxt, our mobile smoking cessation service. Quitxt is delivered through SMS text messaging or Facebook Messenger.

Methods: We collaborated with an advertising agency to conduct in-person recruitment of young adult smokers aged 18-29 years, focusing on urban and rural Spanish-speaking Latino participants, as well as English-speaking rural White and African American participants. Street team members were recruited and trained in a 4-hour session, including a brief introduction to the public health impacts of cigarette smoking and the aims of the project. The street teams made direct, face-to-face contact with smokers in and near smoking areas at 25 bars, clubs, and other venues frequented by young smokers in urban San Antonio and nearby rural areas.

Results: The 3923 interactions by the street teams produced 335 (8.5%) program enrollments. Most participants were English speakers with a mean age of 29.2 (SD 10.6) years and smoked a mean of 8.5 (SD 6.2) cigarettes per day. Among users who responded to questions on gender and ethnicity, 66% (70/106) were women and 56% (60/107) were Hispanic/Latino. Among users ready to make a quit attempt, 22% (17/77) reported 1 tobacco-free day and 16% (10/62) reported maintaining cessation to achieve 1 week without smoking. The response rate to later follow-up questions was low.

Conclusions: Direct outreach in bars and clubs is a useful method for connecting young adult cigarette smokers with mobile cessation services. However, further research is needed to learn more about how mobile services can influence long-term smoking cessation among those recruited through direct outreach, as well as to test the use of incentives in obtaining more useful response rates.

(JMIR Form Res 2022;6(6):e28059) doi:10.2196/28059

KEYWORDS

smoking cessation; young adults; Latinos; mobile intervention; direct recruitment

Introduction

Quitxt (Institute for Health Promotion Research) is a mobile service for smoking cessation designed to guide users through the process of quitting and maintaining abstinence from tobacco use. Quitxt is available via SMS text messaging and Facebook messaging, in both English and Spanish, with enrollment promoted mainly through social media advertising. The program does not offer incentives for responding to assessment questions. Prior studies have demonstrated that Quitxt can be effective in helping up to 1 in 5 users to stop smoking for up to 6 months [1,2]. In this study, we sought to recruit users directly at locations where smokers are frequently encountered: in bars and clubs where alcohol is consumed.

Previous research has found that social media can reach young adults wherever they are, at any time of the day [3]. With the potential for private interactions and public peer outreach, Facebook is a useful, cost-effective recruitment source for young adult smokers [1,4,5]. Mobile devices also have great potential to provide personalized smoking cessation support services, as a Cochrane Review [6] has concluded that text message–based interventions can significantly increase odds of successful cessation of smoking. Many programs have demonstrated that an outreach program using workshops to deliver smoking cessation treatment is effective and easily adoptable by existing public health organizations [7]. Adapting recruitment strategies to expand the program by reaching tobacco users where they are located may help decrease the health-equity gap in tobacco control by improving reach and enrollment of underserved smokers residing in low-income and rural areas who are not reached by traditional cessation services [8].

Despite much research on innovation in cessation services, there were comparatively few reports on methods other than media promotion and advertising for recruiting participants into such services. In a widely cited study, Lando et al [9] reported success from telephone outreach to numbers listed in defined populations for recruiting smokers into telephone counseling. McClure et al [10] found that proactive invitation letters to newsletter-recruited smokers could enhance enrollment in an internet-based service. “Targeting” potential enrollees via direct face-to-face outreach was recommended, but not reported, by Chevalking et al [11] for recruitment to mobile apps similar to the one used in this study. In fact, direct outreach to tobacco users where they can most easily be found is increasingly recognized as an important way to recruit them to cessation services [12-15]. Although internet communication is easy to organize and may be highly cost-effective, person-to-person recruitment to directly reach potential participants in person may be the best way to reach those with lower incomes, lower education levels, and difficulties accessing the internet [13]. Recently published reports described the deployment of “street teams” to reach teens and young adults through peer-to-peer outreach strategies at community venues and events [16]. Outreach at places of employment has been reported for underserved Latino communities at urban construction sites [12]. A very promising approach has been reported from Hong Kong, in which tobacco users are found and recruited directly at outdoor “smoking hotspots” [16-18]. Other very recent research described positive results from the distribution of free nicotine replacement products as a part of direct outreach at smoking hotspots [19-23]. Cigarette smoking and alcohol use are well known to be concomitant behaviors. Most smokers drink alcohol, smokers are more likely to drink than nonsmokers, and those who use both may use them together in the same situations [24,25]. In addition, alcohol consumption is associated with smoking persistence, smoking relapse, and lower odds of quitting success [26-28]. However, there is a lack of research on smokers’ recruitment for mobile cessation services at places where alcohol is consumed, such as bars and clubs. This represents an approach that clearly deserves exploration of feasibility, and thus is the subject of this study.

To explore the feasibility of direct recruitment at venues where alcohol is consumed, we collaborated with an advertising agency (Foundry512) that deployed street teams of outreach workers to recruit smokers to join our Quitxt program at bars, clubs, and other venues in and near San Antonio, Texas. Direct recruitment was done during a brief interval before the COVID-19 pandemic, which made such face-to-face work in bars and clubs unfeasible.

Methods

Ethics Consideration

This study did not require Institutional Review Board approval because it is not a regulated research as defined by the Department of Health and Human Sciences (DHHS) regulations at 45 Code of Federal Regulations (CFR) 46 and FDA regulations at 21 CFR 56 [29].

The proposed program is not funded as research and is not a systematic investigation to test a hypothesis and permit conclusions to be drawn. Further, the purpose is not to investigate the safety or effectiveness of a drug, medical device or biologic.

Target Audience and Recruitment

The target audience for this work was individuals aged 18-29 years, and we focused on urban and rural Spanish-speaking Latino smokers, as well as English-speaking rural White and African American smokers. The face-to-face recruitment was carried out by Foundry512, a full-service advertising agency with extensive experience in the promotion of alcohol products in places where alcohol is consumed.

Selecting and Training Street Team Members

Foundry512 hired and trained 9 part-time workers to form “street teams.” To help ensure street team members could establish a positive relationship with smokers, selection was done to ensure that the team mirrored the culture and demographics of our target audiences. Of the team members, most were bilingual, all had previous experience in outreach marketing, and they

were either former smokers or had smokers in their social circles.

Street teams were trained to use a direct-interaction approach to engage users who were observed using tobacco. The main training objectives were to raise awareness and empathy for the subject and to prepare the teams to follow the program's methodology. The training was conducted in 2 sessions that lasted 2 and 4 hours, respectively. During the first session, street team members received information about smoking, risk factors, side effects, and how the advertising industry has influenced smoking. In addition, the elements that contribute to successful in-person interactions with someone who tries to quit or is thinking about quitting were explained. The second session included a brief introduction to the public health impacts of cigarette smoking and the aims of the Quitxt program and its mobile platforms. Lastly, the training focused on preparing the teams for questions or doubts that smokers may have during the interaction, using empathetic approaches and responses that reflect the challenge of quitting, and learning ways to handle difficult situations or extremely negative responses. Both modeling demonstrations and role playing were used to help the street teams learn the guidelines for interaction, feel more comfortable with the questions to be asked, and make their interactions flow naturally and empathetically.

Approaching Potential Participants

Foundry512 identified and selected 25 bars, clubs, and other venues and events frequented by young adult smokers in urban San Antonio and nearby rural areas. They developed a daily schedule and assigned street team members to different venues and events. When smokers were encountered, street team members guided the interaction using three short questions designed to obtain information about the intention or motivation that smokers had to quit smoking: (1) How important is it to you to quit smoking? (2) How confident are you in your ability to quit smoking? and (3) How ready do you feel to quit smoking? Response options used a rating scale of 1-10 to assess the smokers' level of motivation and readiness to quit smoking. This schema was designed by Foundry512 to be used as a guide to create fluid conversation and to manage complicated interactions with smokers who are apathetic or conversations that would exceed time limits.

The street teams made face-to-face contact with smokers at selected locations and conducted interactions that had an average duration of 4.5 minutes each. In the interactions, street teams first established rapport with the smokers and then toward the end of the engagement introduced the Quitxt mobile cessation program, explained its functions and benefits, and invited them to join the program.

The Quitxt Mobile Cessation Program

Quitxt is a bilingual, evidence-based mobile cessation service that is free to use and culturally appropriate. It turns the participants' phones into a personal coach to help young adults quit smoking. To enroll in the program, participants should be 18 years of age or older, current smokers, and willing to set a quit date within 14 days and provide baseline data [2]. The Quitxt SMS text messaging and Facebook Messenger platforms each contain a complex sequence of interactive messaging, beginning with collection of baseline data that include basic demographics (ie, age, ethnicity, gender), number of cigarettes smoked per day, and e-cigarette use. Participants are then prompted to either choose "quit tomorrow" or set a "quit date" within 2 weeks. Based on the selection, a specific sequence of messages follows. The program provides motivational messages, tips to manage cravings and difficult situations, and 24/7 support. After their quit date, enrollees are also encouraged to text "help" if they are having difficulty avoiding cigarettes. When they text "help," the system texts to ask if the help needed is due to "stress" or "mood"; depending on their text reply, they are then sent either a link to breathing exercises (for stress) or a message with links to diverting, humorous videos (for mood) [2].

Data Analysis

To describe the characteristics of participants referred by the street teams and enrolled in the program, categorical and continuous variables were summarized with frequency distribution, cross-tabulation, mean, and standard deviation as appropriate. Cessation rates were calculated as the proportion of users who reported abstinence among the total active users at the time of the assessment, which were completed at 1, 7, 28, and 72 days from the quit date set by each smoker [30].

Results

The street teams reported 3923 interactions and 353 enrollments of smokers who fit the target group. According to street team members, the most useful question they asked was "How important is it to you to quit smoking?" The least useful was "How ready do you feel to quit smoking?" Street team members stated that this question generated negative reactions and refusals. Therefore, based on the team members' assessments, Foundry512 decided to take out this question with the purpose of maximizing positive reactions from smokers. Interestingly, street team members reported that Latino smokers did not use Spanish as their primary language in these types of social settings. It is also notable that suspicion of the street teams' legitimacy and lack of receptivity were overcome by designing and providing a badge and lanyard for the street teams to look like a cohesive unit with a unified purpose.

Although 353 enrollments were reported by the street teams, data were obtained from 335 users who formally enrolled in the Quitxt SMS (n=317) and Quitxt Facebook Messenger (n=18) services (Table 1).

All but 5 participants preferred the protocol in English. Only 114 participants provided their age (mean 29.2, SD 10.6 years). Of the 106 participants who reported their gender, 66% (n=70) were female. Ethnicity was reported by 107 enrollees and 56% (n=60) identified themselves as Hispanic/Latino. Among the 80 who reported their current tobacco use rate, the mean consumption was 8.5 (SD 6.2) cigarettes per day.

Among these 335 formally enrolled participants, 77 (23%) reported that they were ready to make a quit attempt within 2 weeks. The protocol included questions about smoking cessation during the process of quitting. Among users who were ready to

make a quit attempt, 22% (17/77) reported abstinence at quit day 1, and 16% (10/62) reported maintaining cessation to achieve 1 week without tobacco use. Response rates to later follow-up queries were very low. Only 2 responded to the query about tobacco use 28 days after their quit attempt, and 1 of those reported maintained cessation. There were 3 participants who responded to a similar question 72 days after their quit attempt, and all 3 of these reported that cessation had been maintained for that interval.

Table 1. Characteristics of participants (N=335) recruited by street teams.

Characteristics	Value[a]
Preferred language (N=335), n (%)	
English	330 (98)
Spanish	5 (2)
Age (years; n=114), mean (SD)	29.2 (11)
Gender (n=106), n (%)	
Male	36 (34)
Female	70 (66)
Ethnicity (n=107), n (%)	
Hispanic/Latino	60 (56)
Non-Hispanic/Latino	47 (44)
Cigarettes smoked/day (n=80), mean (SD)	8.5 (6)
Ready to make a quit attempt (N=335), n (%)	77 (23)
Cessation at quit day 1 (n=77)	17 (22)
Cessation at quit day 7 (n=62)	10 (16)

[a]Percentages were calculated based on responses received, as some participants did not respond to or skipped over questions.

Discussion

Principal Findings

Our results suggested that face-to-face recruitment of young adult smokers at venues where alcohol is served (ie, clubs, bars, restaurants) was a feasible strategy to enroll both English- and Spanish-speaking young adults into a mobile cessation service. As in the studies conducted by Asfar et al [12], Chan et al [15], and Saw et al [16], the Quitxt street teams directly recruited participants at places that are frequented by young adult smokers; however, none of their studies involved a mobile phone cessation program. Given that young adults are less likely to seek or use traditional cessation services, such as quitlines [2], direct outreach to young adult smokers by street teams of peers may improve reach and enrollment to smoking cessation services, particularly if the program is tailored to the participants' culture, language, and media use.

There are some limitations to consider. This was not a randomized controlled study and we were unable to assess the effectiveness of the street teams' approach compared with social media advertising, which is the most common strategy used by Quitxt to promote and recruit participants into the program. The fidelity of the intervention was not assessed, and the street team members' adherence to its methodology was unclear. As has happened to many programs across the country, the COVID-19 pandemic impacted our program activities. Face-to-face recruitment by Foundry512 had to be canceled 2 weeks before completing the scheduled activities due to shutdown of bars, restaurants, clubs, and public events to protect the public and avoid the spread of the virus. Even though COVID-19 cut short the face-to-face recruitment efforts, the street teams' approach showed promising results regarding its feasibility in recruiting for a mobile cessation service.

There were lessons learned and challenges experienced during the implementation. First, selecting street team members who share similar demographic characteristics as the target groups is important and facilitates interaction and engagement. Second, implementing hands-on training that incorporates role play is essential to building team confidence. Third, having formal badges, business cards, or other forms of formal identification for street teams provides legitimacy, generates trust, and facilitates interactions with potential participants. The initial lack of standardized identification during the first activation caused people to be less receptive to the street teams, which led potential participants to question the legitimacy of the activity. Foundry512 resolved this issue by designing and providing a badge and lanyard that included the team member's name and the Quitxt and University of Texas Health Science Center at San Antonio logos. In this way, the street team members looked like a formal cohesive unit with a unified purpose. Fourth, having a contingency plan is important, as well as verifying representation of the target groups in the venues selected during planning. During the first activation, the target audiences were not found at the scheduled locations. Additional research was carried out by Foundry512 as part of a contingency plan and new venues were successfully identified by the team. Lastly, maintaining close communication and creating opportunities for additional training is crucial for adapting the intervention

to unexpected situations. The main challenge faced during implementation was the COVID-19 outbreak, which caused both the participants and the street teams to express feelings of concern and vulnerability. As the situation developed day by day, Foundry512 adapted the street team methodology to meet rising concerns and implement safety measures recommended by the local, state, and federal governments. Team members were trained on safety protocols to keep interactions with participants safe (ie, social distancing, wearing masks). Starting with the second activation, street team members were provided with bilingual enrollment cards that included a QR code and 3-step instructions on how to enroll in Quitxt. This provided a safe way to give instructions to interested young smokers, while avoiding close contact with the person or their cell phones.

Conclusion

Findings from this exploratory study showed that direct outreach in bars and clubs is a feasible method and has great potential for enrolling participants into a mobile smoking cessation program. Nearly 4000 tobacco users were approached and 335 (8.5%) formally registered as enrollees in the program. Among these individuals, approximately 1 in 5 were ready to make a quit attempt and approximately 1 in 7 achieved 1 week of nonsmoking. However, subsequent response rates to assessment questions were too low to draw any conclusions about long-term impact on the users' smoking status. We concluded that direct outreach in bars and clubs is a useful method for connecting young adult cigarette smokers with mobile cessation services and the use of incentives could help obtain more useful response rates. However, further research is needed to learn more about how mobile services can influence long-term smoking cessation for users recruited through face-to-face interactions.

Acknowledgments

The project is funded by the Cancer Prevention & Research Institute of Texas (grant PP170099). The authors wish to thank Vivian Cortez, MPH, for her help editing the manuscript.

Conflicts of Interest

None declared.

References

1. Chalela P, McAlister AL, Muñoz E, Despres C, Akopian D, Kaghyan S, et al. Reaching Latinos through social media and SMS for smoking cessation. In: Ramirez AG, Trapido EJ, editors. Advancing the Science of Cancer in Latinos. Cham, Switzerland: Springer International Publishing; 2020:187-196.
2. Ramirez AG, Chalela P, Akopian D, Munoz E, Gallion KJ, Despres C, et al. Text and mobile media smoking cessation service for young adults in South Texas: operation and cost-effectiveness estimation. Health Promot Pract 2017 Jul;18(4):581-585. [doi: 10.1177/1524839917705130] [Medline: 28438055]
3. Free C, Knight R, Robertson S, Whittaker R, Edwards P, Zhou W, et al. Smoking cessation support delivered via mobile phone text messaging (txt2stop): a single-blind, randomised trial. The Lancet 2011 Jul 02;378(9785):49-55 [FREE Full text] [doi: 10.1016/S0140-6736(11)60701-0] [Medline: 21722952]
4. Ramo DE, Rodriguez TMS, Chavez K, Sommer MJ, Prochaska JJ. Facebook recruitment of young adult smokers for a cessation trial: methods, metrics, and lessons learned. Internet Interv 2014 Apr;1(2):58-64 [FREE Full text] [doi: 10.1016/j.invent.2014.05.001] [Medline: 25045624]
5. Thrul J, Klein AB, Ramo DE. Smoking cessation intervention on facebook: which content generates the best engagement? J Med Internet Res 2015 Nov 11;17(11):e244 [FREE Full text] [doi: 10.2196/jmir.4575] [Medline: 26561529]
6. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2016 Apr 10;4:CD006611 [FREE Full text] [doi: 10.1002/14651858.CD006611.pub4] [Medline: 27060875]
7. Selby P, Voci S, Zawertailo L, Baliunas D, Dragonetti R, Hussain S. Public health impact of a novel smoking cessation outreach program in Ontario, Canada. BMC Public Health 2018 Sep 14;18(1):1117 [FREE Full text] [doi: 10.1186/s12889-018-6012-6] [Medline: 30217187]
8. Best practices user guide: health equity in tobacco prevention and control. Centers for Disease Control and Prevention. 2015. URL: https://www.cdc.gov/tobacco/stateandcommunity/best-practices-health-equity/index.htm [accessed 2021-06-30]
9. Lando HA, Hellerstedt WL, Pirie PL, McGovern PG. Brief supportive telephone outreach as a recruitment and intervention strategy for smoking cessation. Am J Public Health 1992 Jan;82(1):41-46. [doi: 10.2105/ajph.82.1.41] [Medline: 1536333]
10. McClure JB, Greene SM, Wiese C, Johnson KE, Alexander G, Strecher V. Interest in an online smoking cessation program and effective recruitment strategies: results from Project Quit. J Med Internet Res 2006 Aug 22;8(3):e14-e11 [FREE Full text] [doi: 10.2196/jmir.8.3.e14] [Medline: 16954124]
11. Chevalking SKL, Ben Allouch S, Brusse-Keizer M, Postel MG, Pieterse ME. Identification of users for a smoking cessation mobile app: quantitative study. J Med Internet Res 2018 Apr 09;20(4):e118 [FREE Full text] [doi: 10.2196/jmir.7606] [Medline: 29631988]

12. Asfar T, Caban-Martinez AJ, McClure LA, Ruano-Herreria EC, Sierra D, Gilford Clark G, et al. A cluster randomized pilot trial of a tailored worksite smoking cessation intervention targeting Hispanic/Latino construction workers: Intervention development and research design. Contemp Clin Trials 2018 Apr;67:47-55 [FREE Full text] [doi: 10.1016/j.cct.2018.02.007] [Medline: 29454141]
13. Greiner Safi A, Reyes C, Jesch E, Steinhardt J, Niederdeppe J, Skurka C, et al. Comparing in person and internet methods to recruit low-SES populations for tobacco control policy research. Soc Sci Med 2019 Dec;242:112597 [FREE Full text] [doi: 10.1016/j.socscimed.2019.112597] [Medline: 31670216]
14. Patten CA, Fu S, Vickerman K, Bock MJ, Nelson D, Zhu S, et al. Support person interventions to increase use of quitline services among racially diverse low-income smokers: A pilot study. Addict Behav Rep 2019 Jun;9:100171 [FREE Full text] [doi: 10.1016/j.abrep.2019.100171] [Medline: 31193750]
15. Chan SSC, Cheung YTD, Wan Z, Wang MP, Lam T. Proactive and brief smoking cessation intervention for smokers at outdoor smoking "hotspots" in Hong Kong. J Cancer Educ 2018 Apr;33(2):365-370. [doi: 10.1007/s13187-016-1085-3] [Medline: 27474113]
16. Saw A, Steltenpohl CN, Bankston-Lee K, Tong EK. A community-based "Street Team" tobacco cessation intervention by and for youth and young adults. J Community Health 2018 Apr;43(2):383-390. [doi: 10.1007/s10900-017-0435-3] [Medline: 29071569]
17. Cheung YTD, Cheung Li WH, Wang MP, Lam TH. Delivery of a nicotine replacement therapy sample at outdoor smoking hotspots for promoting quit attempts: a pilot randomized controlled trial. Nicotine Tob Res 2020 Aug 24;22(9):1468-1475. [doi: 10.1093/ntr/ntz138] [Medline: 31404466]
18. Cheung YTD, Lam TH, Li WHC, Wang MP, Chan SSC. Feasibility, efficacy, and cost analysis of promoting smoking cessation at outdoor smoking "hotspots": a pre-post study. Nicotine Tob Res 2018 Nov 15;20(12):1519-1524. [doi: 10.1093/ntr/ntx147] [Medline: 28655173]
19. Fu SS, van Ryn M, Nelson D, Burgess DJ, Thomas JL, Saul J, et al. Proactive tobacco treatment offering free nicotine replacement therapy and telephone counselling for socioeconomically disadvantaged smokers: a randomised clinical trial. Thorax 2016 May;71(5):446-453 [FREE Full text] [doi: 10.1136/thoraxjnl-2015-207904] [Medline: 26931362]
20. Haas JS, Linder JA, Park ER, Gonzalez I, Rigotti NA, Klinger EV, et al. Proactive tobacco cessation outreach to smokers of low socioeconomic status: a randomized clinical trial. JAMA Intern Med 2015 Feb;175(2):218-226 [FREE Full text] [doi: 10.1001/jamainternmed.2014.6674] [Medline: 25506771]
21. Thao V, Nyman JA, Nelson DB, Joseph AM, Clothier B, Hammett PJ, et al. Cost-effectiveness of population-level proactive tobacco cessation outreach among socio-economically disadvantaged smokers: evaluation of a randomized control trial. Addiction 2019 Dec;114(12):2206-2216 [FREE Full text] [doi: 10.1111/add.14752] [Medline: 31483549]
22. Kruse GR, Park ER, Chang Y, Haberer JE, Abroms LC, Shahid NN, et al. Proactively offered text messages and mailed nicotine replacement therapy for smokers in primary care practices: a pilot randomized trial. Nicotine Tob Res 2020 Aug 24;22(9):1509-1514 [FREE Full text] [doi: 10.1093/ntr/ntaa050] [Medline: 32198520]
23. Cheung YTD, Chan CHH, Ho KS, Tang C, Lau CWH, Li WHC, et al. Effectiveness of nicotine replacement therapy sample at outdoor smoking hotspots for initiating quit attempts and use of smoking cessation services: a protocol for a cluster randomised controlled trial. BMJ Open 2020 Apr 07;10(4):e036339 [FREE Full text] [doi: 10.1136/bmjopen-2019-036339] [Medline: 32269028]
24. Guydish J, Passalacqua E, Pagano A, Martínez C, Le T, Chun J, et al. An international systematic review of smoking prevalence in addiction treatment. Addiction 2016 Feb;111(2):220-230 [FREE Full text] [doi: 10.1111/add.13099] [Medline: 26392127]
25. Shiffman S, Balabanis M. Do drinking and smoking go together? Alcohol Health Res World 1996;20(2):107-110 [FREE Full text] [Medline: 31798093]
26. Cargill BR, Emmons KM, Kahler CW, Brown RA. Relationship among alcohol use, depression, smoking behavior, and motivation to quit smoking with hospitalized smokers. Psychol Addict Behav 2001 Sep;15(3):272-275. [Medline: 11563809]
27. Falk DE, Yi H, Hiller-Sturmhöfel S. An epidemiologic analysis of co-occurring alcohol and tobacco use and disorders: findings from the National Epidemiologic Survey on Alcohol and Related Conditions. Alcohol Res Health 2006;29(3):162-171 [FREE Full text] [Medline: 17373404]
28. Le Strat Y, Ramoz N, Gorwood P. In alcohol-dependent drinkers, what does the presence of nicotine dependence tell us about psychiatric and addictive disorders comorbidity? Alcohol Alcohol 2010;45(2):167-172. [doi: 10.1093/alcalc/agp094] [Medline: 20089545]
29. US Department of Health & Human Services . 45 CFR 46. Regulations, Policy & Guidance. URL: https://www.hhs.gov/ohrp/regulations-and-policy/regulations/45-cfr-46/index.html
30. Brown L, Cai TT, DasGupta A. Interval estimation for a binomial proportion. Statist. Sci 2001 May 1;16(2):101-133. [doi: 10.1214/ss/1009213286]

Abbreviations

DHHS: Department of Health and Human Sciences

CFR: Code of Federal Regulations

Edited by A Mavragani; submitted 18.02.21; peer-reviewed by P Medina-Ramirez, R Dragonetti; comments to author 05.04.21; revised version received 30.06.21; accepted 29.01.22; published 02.06.22.

Please cite as:
Chalela P, McAlister AL, Despres C, Muñoz E, Sukumaran P, Akopian D, Kaghyan S, Trujillo J, Ramirez AG
Direct Outreach in Bars and Clubs to Enroll Cigarette Smokers in Mobile Cessation Services: Exploratory Study
JMIR Form Res 2022;6(6):e28059
URL: https://formative.jmir.org/2022/6/e28059
doi:10.2196/28059
PMID:35653173

Original Paper

An Intervention to Connect Patients With Psychosis and Volunteers via Smartphone (the Phone Pal): Development Study

Mariana Pinto da Costa[1,2,3], MD, MSc, PhD; Phone Pal Advisory Groups[2]

[1]Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom

[2]Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, United Kingdom

[3]Institute of Biomedical Sciences Abel Salazar, University of Porto, Porto, Portugal

Corresponding Author:
Mariana Pinto da Costa, MD, MSc, PhD
Institute of Psychiatry, Psychology & Neuroscience
King's College London
16 De Crespigny
London, SE58AB
United Kingdom
Phone: 44 02078480002
Email: mariana.pintodacosta@gmail.com

Abstract

Background: Intervention development is a critical stage. However, evidence indicates that the substandard reporting of intervention details is widespread.

Objective: This study aimed to provide an overview of the guiding frameworks, methodology, and stages for the design and construction of a new complex intervention—the Phone Pal.

Methods: The intervention development process followed the Medical Research Council framework for developing complex interventions as well as the person-based approach. The intervention was developed following the evidence synthesis of a literature review, a focus group study, and a survey after consultation and input from advisory groups with a range of stakeholders, including patients, volunteers, clinicians, and academics.

Results: The developed logic model outlines the contextual factors, intervention, mechanisms of change, and short- and long-term outcomes. The operationalized intervention required matching 1 patient with 1 volunteer to communicate with each other through a smartphone via SMS text messages, WhatsApp messages or email, and audio or video calls. Each participant was encouraged to communicate with their match at least once per week for a 12-week period using informal conversation.

Conclusions: The systematic process and theoretically sound strategy through which this intervention was developed can provide insights to future researchers on the reality of developing and preparing the operationalization of a digital intervention using multiple components.

(JMIR Form Res 2022;6(6):e35086) doi:10.2196/35086

KEYWORDS

intervention; intervention development; digital mental health; psychosis; severe mental illness; volunteering; volunteer; mental health; mental illness; development; design; user centered design; smartphone; mobile phone; mobile health; mHealth; MRC framework; Medical Research Council framework

Introduction

Complex Interventions

The distinctive notion of a complex intervention has emerged since 2000 [1] through debate focused on the definitions of complex interventions and their components [2]. In particular, there is a need to identify the active ingredients [3], especially in multicomponent psychosocial interventions. It has been emphasized that the description of complex interventions is not enough [4]. In fact, the process of decision-making during research and whether relevant stakeholders are involved and how is often not described.

The 2010 CONSORT (Consolidated Standards of Reporting Trials) statement recommends that authors report interventions with “sufficient details to allow replication” [5]. However, evidence indicates that substandard reporting of intervention

details is widespread [6,7], which may be due to the complexity of many nonpharmacological interventions carried out within a social context [4,8].

It has been recommended that complex interventions should be defined as being formed of parts; these may be material, human, theoretical, social, or procedural in nature and may themselves have subdivisions. These can be further stratified into higher and lower realms that exercise power individually, in combination, or as emergent properties [9]. The intervention *whole* refers to the intervention as a single complete entity distinct from the parts that comprise it; its existence depends on its parts. However, some approaches view the whole as being more than the component parts; thus, the parts are insufficient to explain changes in the intervention's outcomes [10]. Intervention developers routinely elicit the views of target users in a variety of ways [11,12]. However, there is still controversy surrounding how best to do this [13].

The Phone Pal as a Case Study

Volunteering programs appear beneficial and can be encouraged as a means of integrating patients with severe mental illness (SMI) into their communities and promoting their recovery [14]. SMI typically refers to someone with a diagnosis of psychosis of >2 years, duration which impairs their functionality [15].

Social relationships constitute a complex and multidimensional construct with both structural and functional components. With respect to patients with psychosis, there is a body of literature describing their social isolation, low quality of life, low self-esteem, sedentary lifestyle, and nonsecure attachments [16-20]. Regarding volunteers, studies have suggested improvements in their quality of life and changes in their attitudes toward people with mental illness [21].

In the current era of innovation, new models of volunteering may arise using technology. These may encourage the recruitment of new kinds of volunteers and also open the possibility of remote volunteering across large distances [22]. Thus, there is an additional need for remote models that enable volunteer support using technology to connect people with mental illness to others [23].

In an empowerment model in which volunteers take an *active* role as proactive citizens (eg, by supporting either a neighbor or someone remotely), digital volunteering could be an invaluable public health resource for society [24]. Having a Phone Pal as an intervention establishing informal communication about daily life could provide a distinctive form of mental and social support to people with mental illness, improving their mental and physical health [25].

Objectives

The aim of this study was to report the process of development of a complex intervention (ie, the development of a logic model and the planning of its operationalization) as well as to report on the resultant intervention that was ultimately developed—the Phone Pal.

Methods

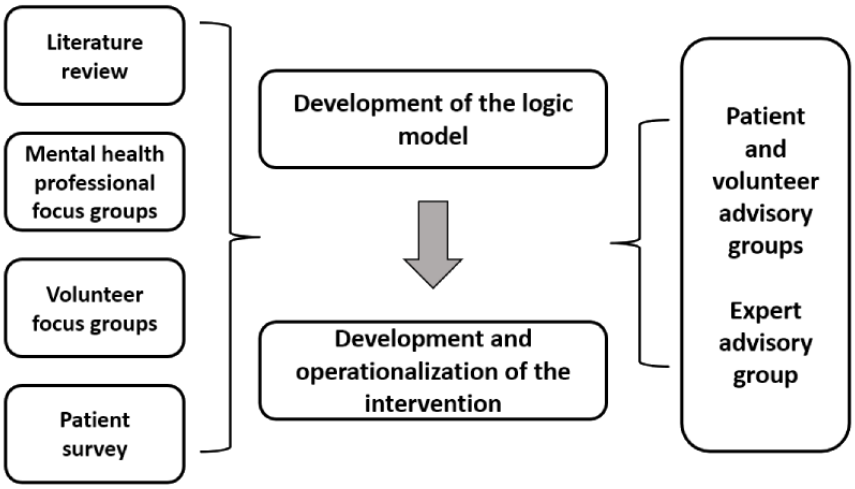
Overview

The process of developing the multicomponent behavioral intervention was performed systematically in different stages. It involved three aspects: (1) developing the logic model, (2) outlining the intervention components, and (3) operationalizing the intervention.

A logic model is a diagrammatic representation of an intervention describing anticipated delivery mechanisms (ie, how resources will be applied to ensure implementation), intervention components (ie, what is to be implemented), hypothesized mechanisms of impact (ie, how an intervention will work), and intended outcomes [26]. Logic models are commonly used to represent the causal processes through which interventions produce outcomes [27].

The development of the Phone Pal logic model followed the recommendations of the 2 guiding frameworks for this study: the UK Medical Research Council (MRC) framework for developing and evaluating complex interventions [28] and the person-based approach [29]. Both frameworks specify the importance of illustrating the theoretical processes that are expected within an intervention and its context. Figure 1 illustrates the various stages of the development pathway followed in this study.

Figure 1. Framework of methods and stages for the intervention development and testing.



Guiding Frameworks

The MRC framework [28] is the most widely used framework for intervention development with multiple components. The MRC framework has four interlinked stages: (1) development, (2) feasibility and piloting, (3) evaluation, and (4) implementation.

This study focuses on the first stage (ie, development), which entails the assessment of the existing evidence base followed by the identification and construction of the theory relating to an intervention. This was relevant to this study, as the mechanisms of volunteering remain unclear.

The person-based approach was selected as a supplementary framework for intervention development [29]. This approach recognizes the specific contextual challenges related to engaging users with digital interventions designed for independent use. This framework argues that interventions must be appealing, easy to use, and relevant to the participants' needs; otherwise, people will not use them.

The person-based approach also advocates that the conceptual modeling described in the MRC framework should consider specific contextual behavioral issues and challenges identified during intervention development. The creation of guiding principles is advised to address contextual challenges and inform an intervention logic model. There are two elements: (1) intervention objectives and (2) core components of the intervention that operationalize the objectives. This approach puts emphasis on component design to improve digital health intervention acceptability and engagement and suggests that all interventions should aim to promote a positive emotional experience [29].

Stages of the Intervention Development

The 4 stages of intervention development used in this study were based on the MRC framework and the person-based approach [28,29] (Multimedia Appendix 1).

Methods Involved in the Intervention Development

Overview

Different methods were used in the intervention development. Some of these activities took place concurrently (eg, the focus group study and survey) and were followed by several iterations of consultations with the advisory groups. The review was conducted between October 2015 and December 2015, the focus groups were conducted between January 2016 and September 2017, the survey was conducted between August 2016 and August 2017, and the advisory groups were consulted from February 2017 to February 2020.

Review of the Literature

A rapid narrative literature review was conducted to enable efficient mapping of the main results related to the specified field. It constitutes a useful technique for intervention development where a broad perspective of the literature is required within a limited timetable [30,31]. Key papers and theories for developing interventions and digital interventions were subsequently searched [32].

Focus Group Study

Stakeholders potentially linked to the provision of volunteering (ie, mental health professionals and volunteers) were interviewed in a focus group format to explore their views regarding the relationship formats between patients and volunteers [32]. A total of 24 focus groups were conducted with 119 participants. The process and findings of the focus groups have been described elsewhere [33,34].

Survey

Patients' preferences with respect to contact with a volunteer were assessed through a survey that evaluated potential relationship formats between patients and volunteers and the patients' preferred volunteer characteristics [22]. A total of 151 patients with psychosis followed in outpatient services in London were interviewed. The findings of this study have been published elsewhere [22].

Volunteer and Patient Advisory Groups

In this research, both patients and volunteers were also included as patient and public involvement advisors [35]. The lead author (MPdC) worked with organizations of volunteers and patients consulting them regarding intervention development [36,37].

The lead author collaborated with a national volunteering association, Befriending Networks [38], which is the umbrella charity for befriending services operating across the United Kingdom and beyond. In September 2017, the lead author organized an event in London that brought several volunteer organizations and volunteers together. The event raised awareness of research on volunteering in mental health and was attended by volunteers involved in charities of mental health. Many of these volunteers contributed to the focus groups conducted during this event. Following the meeting, 2 volunteers also worked with the lead author through the advisory group offering their input for the intervention development process.

To establish patient views, the lead author worked with the Service User and Carer Group Advising on Research at City University of London [39]. This group comprised 14 mental health service users who had personal experience with a variety of mental health conditions, including psychosis.

These groups were consulted throughout the intervention development phase; the meetings involved discussions based on written materials or presentations. The opinions of potential users (ie, both patients and volunteers) were sought concerning which components they thought should be included in an intervention that was aimed toward the provision of opportunities for volunteering in mental health.

The Phone Pal logic model was then developed in consultation with the advisory groups, and it explored a set of person-based intervention components and the potential outcomes that the intervention could affect aligned with the participants' opinions on what to achieve. The groups advised on the suitability of different components and provided recommendations on the role and characteristics of patients and volunteers in this intervention.

These potential users offered important insights to establish the guiding principles of the intervention. They also discussed and specified the key components and objectives that would improve acceptability and engagement with a digital intervention.

Expert Advisory and Consultation Group

Regular meetings with a multidisciplinary team of 30 people comprising principal investigators, postdoctoral researchers, PhD students, researchers, public health experts, anthropologists, psychologists, and psychiatrists at the Unit for Social and Community Psychiatry were used to discuss the intervention development process. The lead author presented the guiding principles, and the team advised on the Phone Pal logic model development.

Evidence Synthesis

The findings emerging from the aforementioned methods were synthesized and presented to the advisory groups. Although there were divergent views on occasions, these group discussions frequently allowed one view to emerge as preferred or, on some matters, there was even consensus from the advisory group members. Further to the consultation with the advisory groups, actions were agreed upon to guide the design of key components of the intervention, the development of the logic model, and the intervention operationalization. The decisions that addressed the identified challenges were acceptable to patients and volunteers, were informed by theory, could be achieved pragmatically within this research, and were endorsed by the expert group.

Ethics Approval

The initial intervention developed was assessed, and changes were requested by the Research Ethics Committee (REC). The final intervention was approved by the REC and the Health Research Authority in the United Kingdom (REC reference 18/EE/0196; protocol 012393).

Results

The Development of the Phone Pal Logic Model

Process for Selecting the Intervention Core Components and Objectives

The process for selecting the intervention core components and objectives was guided by the intervention guiding principles. The guiding principles of the intervention communicate how the objectives and components of the intervention address the particular contextual challenges identified during the process. After an extensive process of gathering the existing evidence, three core intervention components were outlined: (1) a *match* (ie, patient-volunteer), (2) remote communication, and (3) a smartphone, as depicted in Figure 2.

The objectives established were (1) to provide one-to-one access to a *match*, (2) to provide this access to a person remotely, and (3) to allow the communication to occur digitally. Table 1 presents the core components, intervention objectives, and contextual factors.

The following section describes the process of logic model development, portraying the decision-making for the selection of suitable intervention components—namely, the volunteers' characteristics and role, the patients' characteristics and role, and being connected remotely and through a smartphone—using the methods that informed them (Table 2).

Figure 2. Core components of the intervention.

Table 1. Core components, intervention objectives, and contextual factors.

Core components	Intervention objectives	Contextual factors
A "match"	Provide one-to-one access to another person—a match	Patients might be socially isolated
Remote communication	Enable the communication to occur remotely	Patients might face barriers to in-person meetings
A smartphone	Provide access to a person digitally	Technology is integrated into people's lives

Table 2. Components, descriptions, and methods.

Component	Brief description	Literature review	Focus group study	Survey	Patient and volunteer advisory groups	Expert advisory group
Volunteer characteristics	To be inclusive in the definition of volunteers (ie, recruit from a variety of backgrounds and include people with and without personal mental health experience)	✓	✓	✓	✓	
Volunteer role	To provide human contact to patients, establishing informal communication between each other about daily life as part of a more symmetrical relationship	✓	✓	✓	✓	✓
Patient characteristics	To focus on patients with psychosis who are usually the most socially isolated group	✓			✓	✓
Patient role	To obtain human contact from volunteers, establishing informal communication between each other about daily life as part of a more symmetrical relationship	✓	✓		✓	✓
Fully remote	To encourage participants to only communicate with each other remotely	✓				✓
Smartphone	To encourage that participants communicate with each other through a smartphone	✓	✓	✓	✓	

Components

Volunteer Characteristics

Volunteers have varied profiles. It was concluded from systematic reviews [40,41] that there is no *typical* volunteer. These variations [42] encouraged the adoption of a definition of diverse volunteers for the inclusion criteria of this study that enabled volunteer recruitment from a variety of backgrounds.

The volunteer inclusion criteria were widened, recruiting volunteers without experience of mental illness. Although peer support is a theoretical intervention that has been used in web-based psychological interventions to promote co-operation and expertise and reduce loneliness [43], it excludes volunteers who do not possess a shared characteristic.

Volunteers with personal experience of mental illness were also included in this study. Several articles document volunteers who disclose a personal psychiatric history themselves [44-46]. Such volunteers can act as role models and be an inspiration to those with a current mental illness as they are able to demonstrate that *life does go on* and that it is possible to cope with an SMI [44].

In addition, the patient survey demonstrated that most patients (83/148, 56.1%) were interested in a volunteer who had personal experience as a patient in mental health care [22]; these findings added weight to the decision to broaden the inclusion criteria. The focus group study also presented a range of qualities and characteristics related to volunteers. However, it did not advocate for any specific *profile* of volunteer; views ranged from *anyone can be a volunteer* to those who considered that individuals with a *supporting* profile and skills should be the ones providing volunteering support.

For the new intervention, the literature reinforced the decision to have broad inclusion criteria for volunteers, which could encompass a wide range of individuals with or without a possible history of mental issues. The patient and volunteer advisory groups endorsed this choice.

Role of the Volunteer

There are several variations in the potential role of volunteers described in the literature. Although overall these are referred to as *social support* [47], such an approach may comprise emotional, informational, appraisal, and instrumental aspects [48]. Social support has been described as a reciprocal process, where its provision may be as important as its receipt [49].

In the focus group study, stakeholders expressed strong views that the intervention should aim to overcome isolation by connecting patients to additional social contacts, thus using volunteers as *instruments* for *modeling* (ie, as a transition figure).

Of the surveyed patients interested in technology, most (32/56, 57%) thought that the aim of digital volunteering was to make a new friend and had less to do with increasing their activities. They appeared to view technology as a means of establishing contacts with other people and forming friendships. The fact that most of these patients wanted someone as a volunteer with personal experience of mental illness implied a desire to have a more symmetrical relationship.

Some of the members of the expert advisory group initially advised that the volunteers should maintain an *asymmetrical* role of patient support. However, in view of the feedback collected from the focus group study and the survey, most of the expert group members subsequently supported the pursuit of a new model in which a more *symmetrical* relationship was recommended between the 2 matched participants (ie, one that followed the guiding principles of social facilitation, social learning, social role, social comparison, and self-determination).

Encouraging relationship symmetry was a significant decision, distinguishing the current model from more traditional frameworks of befriending or peer support, which exclude

people with or without mental illness, respectively. In this intervention, both were accepted as volunteers. In addition, it was deemed important to incorporate the assessment of social contacts as one of the study outcomes before and after the intervention.

With the endorsement of the patient and volunteer advisory groups, it was decided that the primary role of the volunteer in this intervention would be to offer remote human contact to patients, establishing informal communication between each other about daily life. The volunteer could be a potential role model for the patient, establishing a more symmetrical relationship. Therefore, the focus of this intervention is the patient's and volunteer's social interactions with each other via smartphone as opposed to being centered on being together and engaging in social activities.

Patient Characteristics and Role

The literature describes that, in people with SMI, social isolation has been linked to higher levels of delusions [50], lack of insight [51], and high hospital use [52]. Importantly, previous studies have found a significant association between loneliness and psychosis [53]. The significance of social isolation in patients with psychosis led to the decision to focus this intervention on these individuals.

The focus group study provided insights into the role of patients when outlining the character of the relationships between patients and volunteers.

The survey revealed that younger patients and those with a more recent diagnosis of psychosis were more likely to prefer digital volunteering [22]. However, another study reported that older patients with psychosis were actually more likely to engage with digital interventions [54]. In light of these data, it was opted to include adult patients of all ages in the study. This would also enable the comparison of intervention use according to age. The advisory groups endorsed these proposals for the patients' characteristics and their role in this intervention.

Connected Remotely Through a Smartphone

The concept of volunteers and patients interacting through technology is reinforced in the literature. Technology can help connect people, especially those usually hard to reach [55]. It facilitates more frequent and flexible communication, which is the central element in a relationship. In particular, smartphones are highly portable and in widespread use, presenting new opportunities to deliver interventions, monitor behavior changes, and make communication easier between people who are far away from each other [56-58].

It has been documented that patients with psychosis value technology [59]; are interested in using smartphones [60]; and have been found to engage, adhere, and be satisfied with their use [61-63]. In fact, people with SMI were able to use and adhere to digital interventions [64] even when experiencing negative symptoms [65] and generally found such technology helpful and easy to use. A study evaluating a digital intervention in people with schizophrenia reported that noncompleters were more likely to have severe negative symptoms than completers but found no difference in the incidence of positive or depressive symptoms [66].

People have long established contact with strangers even before technology was in place. This is illustrated by the popularity of *pen pal* letters abroad [67]. This concept supported the decision to design a fully digital intervention and not encourage each match to meet in person during the study.

Among the array of digital tools currently available, smartphones were favored rather than alternatives such as a website or tablet. For the former option, the participants would need constant internet access; with respect to the latter, tablets are commonly larger and less portable than smartphones. The plethora of evidence relating to smartphone use, ownership, and interest in people with psychosis also influenced the decision to deliver the intervention via this modality [59-65]. It was also the most affordable option to enable remote communication between patients and volunteers.

The results from the patient survey revealed that several patients (56/151, 37.1%) were interested in receiving digital volunteering; only a few patients (20/151, 13.2%) did not use technology.

A further important consideration in relation to smartphone use was the advice from the focus group study that veered away from the idea of a separate app. An app designed specifically for people with mental illness to communicate with *healthy* volunteers could potentially risk a further increase in stigma. Instead, a strong recommendation was made to use the normal modalities of communication typically offered by every smartphone (eg, audio calls, video calls, written messages, and emails).

The expert advisory group endorsed the use of this totally remote style of communication to differentiate from face-to-face interactions (Multimedia Appendix 2).

Process for Selecting the Guiding Principles for the Phone Pal Logic Model

Overview

The guiding principles aim to address the contextual challenges that are likely to affect intervention delivery. They consist of 2 aspects (ie, the intervention objectives and the core components used to operationalize them), which were developed to ensure that the intervention met the volunteers' and patients' expectations [68].

This logic model was assembled with the contextual factors and anticipated outcomes following a process of collating existing evidence, varying theories, and conceptualizations. Table 3 illustrates the guiding principles of the intervention, hypothesized mechanisms of change, and outcomes that may be affected by intervention use. The intervention components were selected after feedback to theoretically promote engagement via role modeling and self-efficacy, as further explained.

Table 3. Guiding principles, mechanisms of change, and outcomes.

Guiding principles	Mechanisms of change	Patient outcomes	Volunteer outcomes
Personalization, tailoring, and real-world feel	Engagement	Symptoms and physical activity	Physical activity
Social facilitation, social learning, social role, social comparison, and normative influence	Role model	Social contacts and social comparison	Social distance and social comparison
Self-monitoring, co-operation, and recognition	Self-efficacy	Self-esteem, attachments, and quality of life	Self-esteem and quality of life

Guiding Principles

Different guiding principles were then grouped as drivers of the 3 distinct mechanisms of change [68].

A total of 3 guiding principles were linked with *engagement* as a mechanism of change (ie, *personalization*, *tailoring*, and *real-world feel*). *Personalization* entails the offer of personalized content; *tailoring* encompasses the adaptation of the content to potential needs, interests, personality, context, or other individual factors; and the *real-world feel* emphasizes to the researcher or responsible organization the need to increase the intervention credibility.

In total, 5 guiding principles under the auspice of *role model* form the second grouping (ie, *social facilitation*, *social learning*, *social role*, *social comparison*, and *normative influence*). *Social facilitation* postulates that participants are more likely to perform a target behavior if they are aware that another person is performing the behavior along with them. *Social learning* describes that a participant will be more motivated to perform a target behavior if they are able to observe others performing the behavior. *Social role* suggests that participants will be more likely to perform a target behavior if an intervention adopts a social role. *Social comparison* describes that participants will have a greater motivation to perform a target behavior if they can compare their performance with that of others. *Normative influence* describes that an intervention can provide peer pressure to increase the likelihood that a participant will adopt a target behavior.

The 3 remaining guiding principles were linked with the mechanism of change of *self-efficacy*: *self-monitoring*, *co-operation*, and *recognition*. *Self-monitoring* postulates that an intervention that allows a participant to keep track of their own performance or status supports the participant to achieve their goals. *Co-operation* describes that an intervention can motivate participants to adopt a target behavior by leveraging humans' natural drive to co-operate. *Recognition* indicates that, by offering public recognition for a participant, an intervention can increase the likelihood that the participant will adopt the target behavior.

Mechanisms of Change

The 3 mechanisms of change hypothesized in this model were engagement, role model, and self-efficacy; these are described in the following sections.

Engagement

There are different types of engagement (eg, active or passive), which are accompanied by a heterogeneity of overarching definitions and measurements [69]. The challenge of engagement has been described as one of the main issues with digital interventions [70]. Typically, initial levels of enthusiasm are only maintained for a period after which diminution of use ensues, especially when interventions are used in the *real* world [71]. This may be a problem given that most interventions are designed to become effective over a specific duration. In health care, motivations for compliance can be driven by perceived usefulness and achieving goals; these can facilitate the creation of habits [72].

The behavioral activation theory states that, without specific training, activation consists primarily of the scheduling of pleasant activities [73]. This was key for this intervention. An important objective was that participants were motivated by the intervention and considered it relevant to improving their condition and social context. This study aimed to have the patients engaged with the intervention, interacting with a *real* person—a community volunteer—who would be available to communicate with them and support them.

Role Model

A key underlying theory for this intervention is the weak ties theory, which postulates the importance of less strong ties such as acquaintances to have access to new information outside the core network [74]. A series of principles related to *social support* have been proposed in the Oinas-Kukkonen and Harjumaa [68] framework (ie, social learning, social comparison, and social facilitation, which involve some form of connection with others through the technology). The principles of social learning and social comparison require that the person be aware of what the other person is doing and their progress, potentially achieved by mutual sharing of achievements or witnessing them.

The intervention aimed to provide patients with access to another person (ie, a volunteer) who, through communication, could act as a role model. Theories on *role model* use have advocated that it can be a way of motivating people to perform novel behaviors and inspire them to set ambitious goals [75].

Although the role of volunteers has been described as to *be there* or *do things*, this view may overlap with the conceptualizations of taking a more *passive* or *active* role [14]. In this intervention, it was envisioned that a volunteer would take a somewhat active role realized through being available to communicate with the patient; participating in mutual encouragement to engage in social and physical activities; and aiming to improve their self-esteem, social contacts, and physical activity. The rationale was to motivate participants to choose and communicate through methods that suited them while remaining in contact with their usual friends, thus linking the intervention to their social world.

Self-efficacy

Typically, users of digital interventions must feel motivated and confident to independently use them. Therefore, self-determination theory [76,77] is particularly relevant to understanding how users respond to this intervention and how it differs for patients and volunteers.

Therefore, the intervention components were organized under 3 objectives relevant to the constructs of self-determination theory. This postulates that intrinsic motivation to engage in health behavior change will be enhanced by (1) supporting users' need for autonomy and feeling self-directed; (2) increasing users' sense of competence, control, and confidence; and (3) enhancing users' perceived relatedness or support from the intervention.

Having a chronic mental illness (eg, psychosis) presents several challenges that create a barrier to self-efficacy and affect life quality. Self-esteem has been linked to an intrapersonal influence on self-efficacy, suggesting that increasing self-esteem may subsequently increase self-efficacy [78].

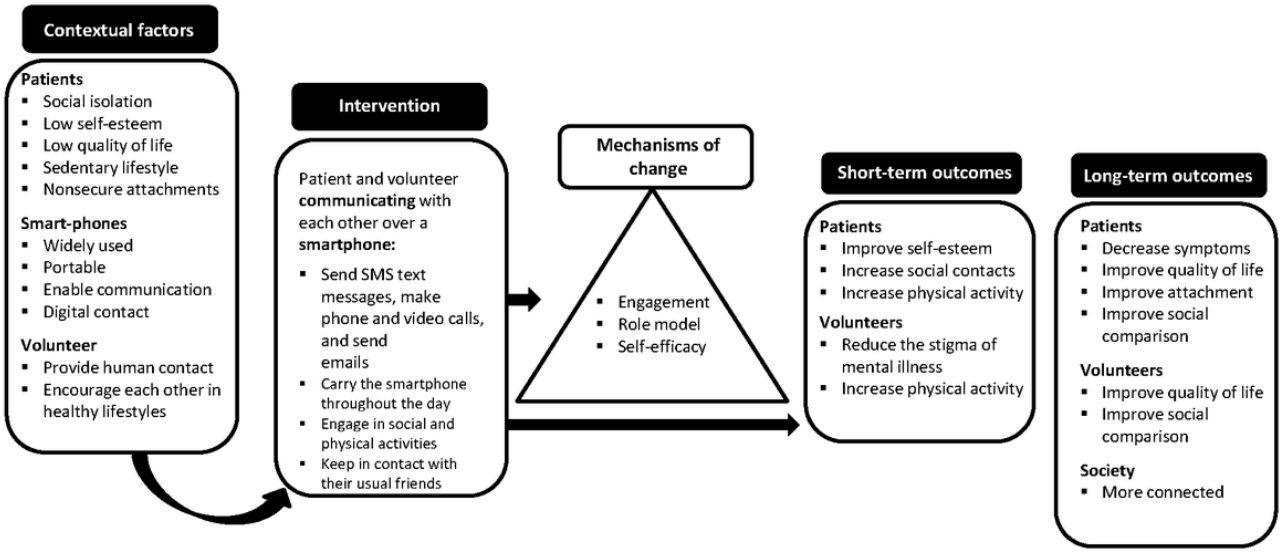
In this study, there were several decisions made to design the intervention with the objective of improving self-efficacy. This entailed the provision of 1 volunteer for each patient with whom they were matched and with whom they could communicate remotely. This allowed them to help their match and to be helped, thus improving their self-esteem and quality of life. Having an attachment to their match could enable a normative influence, co-operation, and recognition. Furthermore, to raise awareness of their physical activity, participants were also encouraged by the researchers to use an app, *Accupedo*, to check their step count, a further means to enhance their self-efficacy [79,80].

Phone Pal Logic Model With Intervention Contextual Factors and Outcomes

Overview

Figure 3 depicts the Phone Pal logic model together with the contextual factors and anticipated outcomes. The model embraces the different components that were selected, the guiding principles promoted, the hypothesized mechanisms of change, and how these concepts are linked to outcomes. This logic model describes the main theoretical pathways connecting the components, principles, mechanisms, and outcomes of this intervention. In reality, the various intervention components may operate on multiple principles and, therefore, multiple mechanisms. An explanation follows for the rationale of why particular components and objectives were included in this intervention.

Figure 3. Logic model with the intervention contextual factors, processes, and outcomes.



Contextual Factors

The Phone Pal logic model encompasses the contextual factors on which the intervention has been based: (1) patients' social isolation, low self-esteem, low quality of life, sedentary lifestyle, and nonsecure attachments; (2) smartphone features (ie, wide availability, portability, and ability to offer digital and remote modes of communication); and (3) community volunteers who provide human contact with others, establishing informal communication about daily life, and offer support, being a role model to patients.

Short-term and Long-term Outcomes

In total, 2 types of outcomes were hypothesized for both patients and volunteers. These were classified according to the time frame (ie, short-term and long-term outcomes).

For patients, potential short-term outcomes were to improve their self-esteem, social contacts, and physical activity. In the longer term, it was felt that additional outcomes could be attained (ie, improving their quality of life, attachment, social comparison, and symptom alleviation).

For volunteers, the short-term outcome goals were to witness a change in their attitudes, with a decrease in their social stigma toward people with mental illness, and to increase their physical activity. Further outcomes could be achieved over a longer

timescale (eg, improvement in quality of life and social comparison).

For both patients and volunteers, it was hypothesized that, through a higher level of engagement, they would become more *activated* and possibly increase their physical activity.

Phone Pal Intervention

The logic model of the planned intervention conceptualized the matching of 1 patient with 1 volunteer to communicate with each other over a smartphone. The core elements were (1) the remote medium for delivery of the intervention, (2) being in contact with another person, and (3) the use of a smartphone.

Developing and Operationalizing the Phone Pal Intervention

Overview

Core components are intrinsically linked with the intervention basis; on a higher level are components that, although less fundamental than the core ones, are still part of the operationalized and envisioned intervention.

The main components describing how the intervention should be and the methods that informed them are summarized in Table 4. The methods and decisions made to operationalize the intervention are summarized in Table 5.

Table 4. Components of the intervention and the methods that informed them.

Component	Brief description	Literature review	Qualitative study	Survey	Patient and volunteer advisory groups	Expert advisory group
Communication duration	To encourage participants to communicate with each other for 12 weeks	✓			✓	
Communication frequency	To encourage participants to communicate with each other at least once per week	✓		✓	✓	
Communication methods	To encourage participants to communicate through the different methods that a smartphone provides (ie, sending messages or emails or making audio or video calls)	✓	✓	✓	✓	✓
Communication content	To encourage participants to have informal communication about daily life		✓		✓	✓

Table 5. Subcomponents of the intervention and the methods that informed them.

Component	Brief description	Literature review	Qualitative study	Survey	Patient and volunteer advisory groups	Expert advisory group
Matching	To match 1 patient with 1 volunteer to communicate with each other during the study	✓			✓	
Volunteer training	To provide training on communication, the boundaries of the relationship, and safeguarding	✓	✓		✓	✓
Patient training	To provide training on communication and the boundaries of the relationship	✓			✓	
Study coordinator	To have a study coordinator available to participants throughout the study offering access to support and supervision				✓	✓
Monitor communication	To monitor the communication of the participants through the offered smartphone				✓	✓
Monitor physical activity	To monitor the step count of the participants through an app on the offered smartphone				✓	✓

Duration, Frequency, and Methods of Communication

For pragmatic reasons, a period of 3 months was chosen in line with several publications reporting interventions with a duration of 12 weeks [54,81-86]. A review identified that the level of commitment should be for a minimum of 4 hours per month [44].

The idea of the communication being tailored by each participant is deemed important to create a persuasive system [68]. Therefore, in this intervention, it was decided to leave it up to each pair to decide which communication modalities they would like to use.

In young adults with a first psychotic episode, it was noted that they preferred a combination of technologies through which to receive mental health care (eg, SMS text message, video, and audio). Among these options, SMS text messages were preferred [87].

Surveyed patients [22] proposed that contacts should be open-ended and weekly and favored SMS text messages followed by email, Skype, Facebook, and audio calls. The

modality *audio calls*, which was not given as a separate option in the survey, was named by different patients when selecting the choice *other*. The methods of communication initially selected for this intervention (ie, audio calls, video calls, SMS text messages, Facebook messages, WhatsApp messages, and emails) were determined according to the results of this survey and the literature. However, Facebook messages were removed following the instructions of the REC.

It was decided to encourage participants to communicate with each other at least once per week. The participants were informed that one of the study aims was to assess how much or how little each pairing decided to communicate. The focus group study findings indicated that people varied in their attitudes toward specific communication methods. In line with the person-based approach, it was decided not to be prescriptive in this regard.

A pragmatic decision was made to give the participants full access to and choice of all possible intervention components (ie, they could choose from sending SMS text messages, WhatsApp messages, or emails or making audio and video calls). The participants could opt to interact with their match through a particular tool, and their paired match could then choose to reciprocate through the same tool or select another modality. Relinquishing control of communication modality use has been supported by previous research [88]. The lead author encouraged volunteers to take the initiative to call patients in the first instance.

The expert advisory group supported the choice of communication method as all the tools were commonly accessible on any smartphone.

While developing the intervention, the lead author was prescriptive in the duration and frequency of the communication but not in the communication modalities. In the process of operationalizing the intervention, the lead author embraced the challenge of how much or how little guidance to provide when researching a psychosocial intervention. In terms of guidance for intervention adherence and use, there can be tension between supporting autonomy while still providing clear guidance on how participants can best change their behavior. Some studies reported concerns that offering too many options can be overwhelming [89]. Bestowing complete control can result in lower intervention use than *tunnelling* core intervention content [90]. Some researchers suggest that tunneling is less overwhelming than allowing patients free choice [68], and it is often used in mental health interventions [91]. However, a *tunnelled* approach [68] in which participants are led sequentially through intervention content, usually in a predefined order or based on a needs assessment, might not always be suitable and was deemed as too imposing. In fact, the person-based approach recommends that, in general, digital interventions should aim to promote user autonomy and offer choices where possible [29].

Communication Content

In this study, communication was between 2 humans and mediated through technology—the smartphone.

The focus groups provided insights into possible suggested conversation content for this intervention. The participants were instructed to establish informal communication about daily life.

Matching Patients and Volunteers

Overview

With respect to matching, previous research on telephone peer support has matched dyads randomly [49]; it was not possible to predict which dyad members would become friends based on interviewer impressions or similarity of individual characteristics. In the Volunteering in Mental Health befriending face-to-face trial, each patient-volunteer pairing was matched based on the instincts of the volunteer coordinator [92]. In the Phone Pal study intervention, it was decided to match on a first-come, first-served basis; the patient and volunteer advisory groups agreed. Specifically, personal characteristics or preferred communication methods were not used to influence this process.

Training Volunteers

The literature reports that volunteer training is compulsory in most programs [44,46,93-96], although some volunteers receive no training [97]. What this potentially means for volunteers varies, from not sharing personal contact details or information to, alternatively, sharing personal information and introducing the patient to friends or family members [96]. Examples of topics covered include expectations and responsibilities of a volunteer, preparation for managing initial meetings, general listening skills, boundaries and guidelines, mental illness, stigma, major diagnoses and symptoms, and conflict management [44,46]. The training offered to volunteers was primarily focused on communication and managing boundaries rather than presenting surplus information about the illness or treatment. The latter was important to avoid the concerns of *medicalizing* or providing unnecessary clinical information, which could jeopardize the establishment of a friendly relationship.

The focus group study provided a variety of views on the advantages and disadvantages of training. Although it was deemed important that training should be provided, there was no clear consensus; some thought it unnecessary. Concerns were also expressed about potentially high expectations, overinvolvement or risk of professionalizing volunteers, and breaching confidentiality. It was advised that these issues should be addressed with adequate training and also by asking volunteers to sign a confidentiality agreement, a common practice in many volunteer programs and research studies. This covered the requirement of volunteers not to disclose information about their matched patient to third-party agencies, friends or family, or anyone outside the research team without consent from the matched patient. In addition, they were requested not to disclose unnecessary information about their volunteer role to family, friends, or colleagues. Breaching confidentiality would only be permitted if there was a risk of serious harm (eg, if a patient expressed planned criminal intent against any individual or intent with a plan to commit suicide or self-harm; if a patient was judged to be at risk of sexual, emotional, or physical abuse; or where not acting on information would increase physical or emotional risk).

Training Patients

Upon the request of the REC, customized training materials were developed for patients; these were adapted from the materials developed for volunteers. The lead author followed existing best practices to maximize the accessibility, usability, and credibility of the intervention for a wide range of people, including those with lower levels of literacy or cognitive impairments [98]. Short sentences, lists, and visual formats were used, and the training interaction was tailored to each patient where appropriate.

Role of the Study Coordinator

The literature indicated that this study had the potential to generate negative feelings in the participants. The sources could be the wait, delay or expectation that their *match* would make contact, or difficulties encountered in dealing with the end of the study or the end of the relationship. Some organizations provide support throughout to volunteers face to face or by using technology while covering users from a wide age range [99]. Offered volunteer supervision has been described in the form of monthly multidisciplinary meetings, one-to-one supervision sessions, or telephone support [46,95-97,100]. In a previous study, when an intervention was remote, web-based support from a staff member and occasional telephone calls were found to be essential for participants to remain in the study and continue to use the digital intervention [101].

Therefore, it was decided that a study coordinator would be available for the participants, providing one-to-one access to support and supervision whenever the participants felt it was required in addition to proactively contacting all participants once per month. This would enable the provision of constant support to volunteers, who are community laypeople who could end up facing potentially unfamiliar situations that they may not know how to deal with (eg, managing patients' behavior or difficulties with the relationship ending). Patients equally require support, although negative symptoms could make them less proactive in contacting the study coordinator. Therefore, the lead author carried an additional work phone. The number was given to all participants, and she could be reached at any time (ie, 24 hours a day and 7 days a week). The monthly contact would routinely be through an audio call; if the participants did not respond, a message would be sent asking for their availability. Although evidence suggests that reminders can improve engagement with digital interventions, there is as yet insufficient data to indicate what types of messages are most likely to promote adherence [102-104]. It also remains unclear how people will respond to these motivational messages. What one person sees as encouraging might demotivate another [105,106].

Monitoring Communication

To improve the understanding of the interactions between patients and volunteers, a practical decision was made to monitor the communication between each pair.

Although various members of the expert advisory group initially raised ethical questions about privacy, most subsequently concurred with this choice given that it offered an additional way to ensure that the content was appropriate and did not raise any safeguarding concerns.

The option of creating a technical *app* for communication monitoring was initially explored. After seeking quotations from different companies, it was established that a purpose-built app would be too costly. An existing app was identified that monitored the content of written communication (SMS text messages, WhatsApp messages, and emails) and the frequency and duration of audio and video calls. Through discussions with the expert advisory group, this app was purchased and used.

Monitoring Physical Activity Through Step Count

In line with the overall aim of the intervention to facilitate improvements in the participants' mental and physical health by establishing informal communication about daily life and encouraging healthy lifestyles [107], physical activity changes were assessed in 2 ways. First, an existing app to monitor step count was installed on the smartphones. Using this, the participants could check their step count and gain awareness of their physical activity. In addition, to address potential issues of participants not always carrying the smartphone, the International Physical Activity Questionnaire [108] was administered both at the beginning and end of the study.

Discussion

Principal Findings

This study reports an innovative model of intervention development using a combined approach with the MRC framework and the person-based approach, which might be followed by other researchers developing interventions.

The Phone Pal logic model was rooted in the weak ties, behavioral activation, and self-determination theories and based on 11 guiding principles and 3 mechanisms of change (ie, engagement, role model, and self-efficacy).

The three core components linked to the logic model are (1) having a match, (2) communicating remotely, and (3) using a smartphone. Each of these components consists of parts and requires characteristics to shape its existence. The characterization of the intervention components is made up of components as well and comprises (1) volunteer characteristics, (2) patient characteristics, (3) role of the volunteer, (4) role of the patient, and (5) remote connection only through a smartphone.

For the intervention development, additional components of duration, frequency, methods, and content of communication were chosen. Finally, for the operationalization of the intervention, the following components were added: matching, training, providing a study coordinator, and monitoring the participants' communication and physical activity.

Strengths and Limitations

A main strength of this intervention development was its systematic process and consultation involving multiple methods and experts. This addresses the literature requirement for a greater focus on the developmental stage of interventions, allowing them to adapt on implementation [109,110]. It was

also informed by relevant frameworks; that is, the MRC framework [28] and the person-based approach [29]. Both endorse the rigorous execution of intervention development and recommend the use of both quantitative and qualitative approaches, encouraging the lead author to combine existing evidence and use the most appropriate methods. This is of particular importance, given the increased awareness and encouragement of the publication of the development phase of interventions [111-114]. There is a notable gap in the literature relating to the systematic design of volunteer interventions despite the research conducted in this area.

Another advantage of this approach is that the intervention development involved participant perspectives throughout in the form of advisors to the research process and as study participants. This adheres to the literature recommendations that suggest that this process ensures the success of digital interventions [115,116]. Furthermore, both frameworks also emphasize the importance of collaborating with experts and clarifying how this involvement influenced the development process [117]. This logic model has been shaped following an extensive process of evidence collation and consultation with numerous stakeholders. It is also simple enough to capture the core elements necessary for this feasibility stage.

Although this approach has several strengths, it also has limitations.

The first limitation relates to the relatively linear process of intervention development in which one process informed the next (ie, a series of studies provided information for the logic model, which subsequently influenced the intervention development and operationalization).

A further potential problem with using the volunteer and patient advisory groups is that they may try to anticipate the needs of others, which they may not do well, rather than simply reporting their own experiences and views [29].

There are also potential challenges that may be faced when developing an intervention that targets different users simultaneously albeit with patients as the primary focus. Patients and volunteers vary in their characteristics. The inclusion criteria are necessarily dissimilar, and the opinions of the 2 parties may disagree on their vision of the intervention and, therefore, influence the other's role or adherence. They may also have different beliefs about their own and others' mental health condition or how they feel about being in contact with others. Not only may these views vary according to their role in the intervention but they may also reflect their individual characteristics (ie, age, cultural background, health condition, digital literacy, or previous experiences). To address this, the lead author spoke with a diverse range of potential users to ensure that she had insight into as many different and relevant perspectives as possible, thus enabling the intervention to be customized for all individuals.

Although the MRC framework is the most widely used approach to intervention development, it only provides a guideline for the relevant elements and the research questions to consider. It does not provide a prescriptive approach to development methodologies related to the varying contexts in which interventions can be used. In addition, there are challenges in incorporating the user perspective. The person-based approach is not intended to replace but to complement the well-known *theory-based* and *evidence-based* approaches that incorporate behavioral science into intervention development [28,118].

A potential limitation can be rooted in the intervention complexity (ie, there are numerous different types of interactions between components, a range of behaviors required by those delivering and receiving the intervention, varied targeted levels, a variety of outcomes, and a high level of flexibility and customization of the intervention [119]). It is challenging for a logic model to accurately capture this complexity. Logic models can be used to model complex interventions that adapt to context. However, it has been suggested that more flexible and dynamic models are required [27].

In addition, for clarity, the logic model portrays the main hypothesized pathways between the components, principles, mechanisms, and outcomes of the intervention. In reality, as previously acknowledged, it is likely that the intervention components operate on multiple principles and, therefore, multiple mechanisms.

Finally, as in other psychosocial interventions, the researchers had little control on how the intervention was delivered and would rely on the participants' descriptions to assess its implementation.

Implications for Future Research

The systematic process and theoretically sound strategy through which the intervention was developed enables other researchers to see clearly how the intervention was designed and why the components were selected in line with the contextual factors that were hypothesized as influential on the necessity and acceptability of the intervention. The logic model developed in this research is relevant to the particular contextual issues present in an intervention for people with psychosis based in London. However, this could provide future researchers with a platform for adapting this to translate to other populations and settings.

The Phone Pal logic model is hypothetical and depicts how core intervention components relate to the intervention principles and hypothesized mechanisms of change. It may not be possible to determine the exact components necessary or those that might have the most impact on the outcomes. Similarly, the mechanisms of change are not fully defined. This model might be built in future research to elucidate how and why any changes might be achieved. Future research should investigate which components contribute most or are essential to a successful intervention (eg, the remote contribution with the volunteer or having a smartphone), which components are superfluous, and how the components interact to influence outcomes [28]. Although this model recognizes the potential contextual factors affecting the acceptability and necessity of this intervention, it may be that in other contexts, different challenges may emerge, requiring the adaptation of this model and modification or substitution of the current components. Future research should explore how these components operate, which principles they promote, and the associated mechanisms. It may be plausible

that other components could usefully contribute. Studies rarely specify whether the factors related to the delivery context affect which components are delivered, but the factors themselves are not always easy to capture. Whether these principles are useful targets for interventions for people with other conditions or in other settings is another topic for further investigation.

Previous logic models on volunteering [120,121] have hypothesized how volunteering may operate, focusing on the organization and requirements such as the acquisition of skills and retention of volunteers. In contrast, the logic model proposed in this study is the first to outline and place a focus on the hypothesized relationships between the patient and volunteer engaging in digital and remote communication. This may be helpful for the wider conceptualization of volunteering, distancing itself from the principle of the *organization* and focusing on person-to-person interactions. Previously, no specific model has been conceptualized that hypothesizes the relationships between patients and volunteers either face to face or digitally. This may inspire future research to potentially construct a model that focuses on face-to-face interactions.

The operationalization of this intervention could inform future research as it provides a clear map of the processes and outcomes of the development of a digital intervention using different components.

The main implication of this intervention development is allowing for further testing in a feasibility study before the final stages of the MRC framework—evaluation and implementation.

Conclusions

The intervention development followed the MRC framework for developing complex interventions as well as the person-based approach. The intervention was developed following the evidence synthesis of a literature review, a focus group study, and a survey after consultation with advisory groups and input from a range of stakeholders, including patients, volunteers, clinicians, and academics. The developed logic model outlines the contextual factors, intervention, and short- and long-term outcomes. The operationalized intervention required matching 1 patient with 1 volunteer to communicate with each other through a smartphone via SMS text messages, WhatsApp messages, emails, and audio or video calls. All participants were encouraged to carry the smartphone throughout the day, engage in social and physical activities, and keep in contact with their usual friends. Each participant was encouraged to communicate with their match at least once per week for a 12-week period using informal conversation. The intervention provides guidance for the duration and frequency of communication (ie, once per week) but does not provide any recommendation regarding the different communication methods (ie, SMS text messages, WhatsApp messages, emails, and audio and video calls), leaving this decision completely to each patient-volunteer pair.

The findings of this study can inform future research. It presents the initial development phases of the intervention, the map of processes and outcomes from the logic model, and the actual operationalization. Practical issues faced could provide useful insights into the reality of preparing the operationalization of a digital intervention using multiple components.

Acknowledgments

This research has been supported by the National Institute for Health Research Collaboration for Leadership in Applied Health Research and Care North Thames at Barts Health National Health Service Trust and the East London National Health Service Foundation Trust.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Overview of the 4 stages of the intervention development process using an adapted version of the Medical Research Council framework and the person-based approach.
[PDF File (Adobe PDF File), 599 KB - formative_v6i6e35086_app1.pdf]

Multimedia Appendix 2
Smartphone provided during the study.
[PDF File (Adobe PDF File), 392 KB - formative_v6i6e35086_app2.pdf]

References

1. A framework for the development and evaluation of RCTs for complex interventions to improve health. Medical Research Council. 2000 Apr. URL: https://www.ntnu.no/wiki/download/attachments/53514442/A%20Framework%20for%20development%20and%20evaluation%20of%20RCTs%20for%20Complex%20Interventions%20to%20Improve%20Health.pdf?version=1&modificationDate=1348243868000&api=v2 [accessed 2022-05-16]

2. Davidson KW, Goldstein M, Kaplan RM, Kaufmann PG, Knatterud GL, Orleans CT, et al. Evidence-based behavioral medicine: what is it and how do we achieve it? Ann Behav Med 2003 Dec;26(3):161-171. [doi: 10.1207/S15324796ABM2603_01] [Medline: 14644692]
3. Sutcliffe K, Thomas J, Stokes G, Hinds K, Bangpan M. Intervention Component Analysis (ICA): a pragmatic approach for identifying the critical features of complex interventions. Syst Rev 2015 Oct 29;4:140 [FREE Full text] [doi: 10.1186/s13643-015-0126-z] [Medline: 26514644]
4. Wells M, Williams B, Treweek S, Coyle J, Taylor J. Intervention description is not enough: evidence from an in-depth multiple case study on the untold role and impact of context in randomised controlled trials of seven complex interventions. Trials 2012 Jun 28;13:95 [FREE Full text] [doi: 10.1186/1745-6215-13-95] [Medline: 22742939]
5. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010 Mar 23;340:c332 [FREE Full text] [doi: 10.1136/bmj.c332] [Medline: 20332509]
6. Pino C, Boutron I, Ravaud P. Inadequate description of educational interventions in ongoing randomized controlled trials. Trials 2012 May 18;13:63 [FREE Full text] [doi: 10.1186/1745-6215-13-63] [Medline: 22607344]
7. Hoffmann TC, Erueti C, Glasziou PP. Poor description of non-pharmacological interventions: analysis of consecutive sample of randomised trials. BMJ 2013 Sep 10;347:f3755 [FREE Full text] [doi: 10.1136/bmj.f3755] [Medline: 24021722]
8. Hooper R, Froud RJ, Bremner SA, Perera R, Eldridge S. Cascade diagrams for depicting complex interventions in randomised trials. BMJ 2013 Nov 15;347:f6681. [doi: 10.1136/bmj.f6681] [Medline: 24243975]
9. Clark AM. What are the components of complex interventions in healthcare? Theorizing approaches to parts, powers and the whole intervention. Soc Sci Med 2013 Sep;93:185-193. [doi: 10.1016/j.socscimed.2012.03.035] [Medline: 22580076]
10. Elder-Vass D. The Causal Power of Social Structures Emergence: Structure and Agency. Cambridge, UK: Cambridge University Press; 2010.
11. Yen PY, Bakken S. Review of health information technology usability study methodologies. J Am Med Inform Assoc 2012;19(3):413-422 [FREE Full text] [doi: 10.1136/amiajnl-2010-000020] [Medline: 21828224]
12. Van Velsen L, Wentzel J, Van Gemert-Pijnen JE. Designing eHealth that matters via a multidisciplinary requirements development approach. JMIR Res Protoc 2013 Jun 24;2(1):e21 [FREE Full text] [doi: 10.2196/resprot.2547] [Medline: 23796508]
13. Iivari J, Isomäki H, Pekkola S. The user – the great unknown of systems development: reasons, forms, challenges, experiences and intellectual contributions of user involvement. Inf Syst J 2010 Mar;20(2):109-117. [doi: 10.1111/j.1365-2575.2009.00336.x]
14. Pinto da Costa M, Foster R, Gillard S, Priebe S. Volunteering in mental health. In: Okpaku SO, editor. Innovations in Global Mental Health. Cham, Switzerland: Springer; 2020:1-28. [doi: 10.1007/978-3-030-57296-9_72]
15. Ruggeri M, Leese M, Thornicroft G, Bisoffi G, Tansella M. Definition and prevalence of severe and persistent mental illness. Br J Psychiatry 2000 Aug;177:149-155. [doi: 10.1192/bjp.177.2.149] [Medline: 11026955]
16. Rotondi V, Stanca L, Tomasuolo M. Connecting alone: smartphone use, quality of social interactions and well-being. J Econ Psychol 2017 Dec;63:17-26. [doi: 10.1016/j.joep.2017.09.001]
17. Stuckey MI, Carter SW, Knight E. The role of smartphones in encouraging physical activity in adults. Int J Gen Med 2017 Sep 12;10:293-303 [FREE Full text] [doi: 10.2147/IJGM.S134095] [Medline: 28979157]
18. Bedrov A, Bulaj G. Improving self-esteem with motivational quotes: opportunities for digital health technologies for people with chronic disorders. Front Psychol 2018 Nov 2;9:2126 [FREE Full text] [doi: 10.3389/fpsyg.2018.02126] [Medline: 30450071]
19. Thomée S. Mobile phone use and mental health. A review of the research that takes a psychological perspective on exposure. Int J Environ Res Public Health 2018 Nov 29;15(12):2692 [FREE Full text] [doi: 10.3390/ijerph15122692] [Medline: 30501032]
20. Aledavood T, Torous J, Triana Hoyos AM, Naslund JA, Onnela JP, Keshavan M. Smartphone-based tracking of sleep in depression, anxiety, and psychotic disorders. Curr Psychiatry Rep 2019 Jun 04;21(7):49 [FREE Full text] [doi: 10.1007/s11920-019-1043-y] [Medline: 31161412]
21. Abayomi O, Adelufosi A, Olajide A. Changing attitude to mental illness among community mental health volunteers in south-western Nigeria. Int J Soc Psychiatry 2013 Sep;59(6):609-612. [doi: 10.1177/0020764012448781] [Medline: 22781779]
22. Pinto da Costa M, Chevalier A, Farreny A, Cassidy M, Leverton M, Toner S, et al. How would patients with psychosis like to be in contact with a volunteer: face-to-face or digitally? PLoS One 2019 May 16;14(5):e0216929 [FREE Full text] [doi: 10.1371/journal.pone.0216929] [Medline: 31095611]
23. Pinto da Costa M. Can social isolation caused by physical distance in people with psychosis be overcome through a Phone Pal? Eur Psychiatry 2020 May 22;63(1):e61 [FREE Full text] [doi: 10.1192/j.eurpsy.2020.53] [Medline: 32441240]
24. User empowerment in mental health – a statement by the WHO Regional Office for Europe. World Health Organization. 2010. URL: https://www.euro.who.int/__data/assets/pdf_file/0020/113834/E93430.pdf [accessed 2022-05-16]
25. Pinto da Costa M, Phone Pal Advisory Groups. Volunteering via smart-phone for people with psychosis-protocol of a feasibility trial. Front Psychiatry 2021 Nov 30;12:742202 [FREE Full text] [doi: 10.3389/fpsyt.2021.742202] [Medline: 34916970]

26. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Council guidance. BMJ 2015 Mar 19;350:h1258 [FREE Full text] [doi: 10.1136/bmj.h1258] [Medline: 25791983]
27. Mills T, Lawton R, Sheard L. Advancing complexity science in healthcare research: the logic of logic models. BMC Med Res Methodol 2019 Mar 12;19(1):55 [FREE Full text] [doi: 10.1186/s12874-019-0701-4] [Medline: 30871474]
28. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008 Sep 29;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
29. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
30. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
31. Khangura S, Polisena J, Clifford TJ, Farrah K, Kamel C. Rapid review: an emerging approach to evidence synthesis in health technology assessment. Int J Technol Assess Health Care 2014 Jan;30(1):20-27. [doi: 10.1017/S0266462313000664] [Medline: 24451157]
32. Pinto da Costa M. Volunteering in mental health - towards a new model of support, in Barts and the London School of Medicine and Dentistry. Queen Mary University of London. 2020.
33. Pinto da Costa M, Conneely M, Monteiro da Silva F, Toner S. Stakeholders' views on volunteering in mental health: an international focus group study. BMJ Open 2022 Mar 17;12(3):e052185 [FREE Full text] [doi: 10.1136/bmjopen-2021-052185] [Medline: 35301203]
34. Pinto da Costa M. Conducting cross-cultural, multi-lingual and multi-country focus groups: guidance for researchers. Int J Qual Methods 2021 Nov 05;20:160940692110499. [doi: 10.1177/16094069211049929]
35. Hoddinott P, Pollock A, O'Cathain A, Boyer I, Taylor J, MacDonald C, et al. How to incorporate patient and public perspectives into the design and conduct of research. F1000Res 2018 Jun 18;7:752 [FREE Full text] [doi: 10.12688/f1000research.15162.1] [Medline: 30364075]
36. Manafò E, Petermann L, Vandall-Walker V, Mason-Lai P. Patient and public engagement in priority setting: a systematic rapid review of the literature. PLoS One 2018 Mar 2;13(3):e0193579 [FREE Full text] [doi: 10.1371/journal.pone.0193579] [Medline: 29499043]
37. Goldsmith LP, Morshead R, McWilliam C, Forbes G, Ussher M, Simpson A, et al. Co-producing randomized controlled trials: how do we work together? Front Sociol 2019 Mar 29;4:21 [FREE Full text] [doi: 10.3389/fsoc.2019.00021] [Medline: 33869347]
38. Befriending Networks. URL: https://www.befriending.co.uk/ [accessed 2022-05-17]
39. Simpson A, Jones J, Barlow S, Cox L, Service User and Carer Group Advising on Research (SUGAR). Adding SUGAR: service user and carer collaboration in mental health nursing research. J Psychosoc Nurs Ment Health Serv 2014 Jan;52(1):22-30. [doi: 10.3928/02793695-20131126-04] [Medline: 24305906]
40. Hallett C, Klug G, Lauber C, Priebe S. Volunteering in the care of people with severe mental illness: a systematic review. BMC Psychiatry 2012 Dec 13;12:226 [FREE Full text] [doi: 10.1186/1471-244X-12-226] [Medline: 23237048]
41. Toner S, Hickling LM, Pinto da Costa M, Cassidy M, Priebe S. Characteristics, motivations and experiences of volunteer befrienders for people with mental illness: a systematic review and narrative synthesis. BMC Psychiatry 2018 Dec 04;18(1):378 [FREE Full text] [doi: 10.1186/s12888-018-1960-z] [Medline: 30514260]
42. Bussell H, Forbes D. Understanding the volunteer market: the what, where, who and why of volunteering. Int J Nonprofit Volunt Sect Mark 2002 Sep;7(3):244-257. [doi: 10.1002/nvsm.183]
43. Naslund JA, Aschbrenner KA, Marsch LA, Bartels SJ. The future of mental health care: peer-to-peer support and social media. Epidemiol Psychiatr Sci 2016 Apr;25(2):113-122 [FREE Full text] [doi: 10.1017/S2045796015001067] [Medline: 26744309]
44. McCorkle BH, Dunn EC, Wan YM, Gagne C. Compeer friends: a qualitative study of a volunteer friendship programme for people with serious mental illness. Int J Soc Psychiatry 2009 Jul;55(4):291-305. [doi: 10.1177/0020764008097090] [Medline: 19553360]
45. Mitchell G, Pistrang N. Befriending for mental health problems: processes of helping. Psychol Psychother 2011 Jun;84(2):151-169. [doi: 10.1348/147608310X508566] [Medline: 22903854]
46. Copello AG, Velleman RD, Howling VM. The use of volunteers as befrienders within a community alcohol team. J Subst Misuse 2009 Jul 12;3(4):189-199. [doi: 10.3109/14659899809053501]
47. Faulkner M, Davies S. Social support in the healthcare setting: the role of volunteers. Health Soc Care Community 2005 Jan;13(1):38-45. [doi: 10.1111/j.1365-2524.2005.00526.x] [Medline: 15717905]
48. Langford CP, Bowsher J, Maloney JP, Lillis PP. Social support: a conceptual analysis. J Adv Nurs 1997 Jan;25(1):95-100. [doi: 10.1046/j.1365-2648.1997.1997025095.x] [Medline: 9004016]

49. Heller K, Thompson MG, Trueba PE, Hogg JR, Vlachos-Weber I. Peer support telephone dyads for elderly women: was this the wrong intervention? Am J Community Psychol 1991 Feb;19(1):53-74. [doi: 10.1007/BF00942253] [Medline: 1867151]
50. Garety PA, Kuipers E, Fowler D, Freeman D, Bebbington PE. A cognitive model of the positive symptoms of psychosis. Psychol Med 2001 Feb;31(2):189-195. [doi: 10.1017/s0033291701003312] [Medline: 11232907]
51. White R, Bebbington P, Pearson J, Johnson S, Ellis D. The social context of insight in schizophrenia. Soc Psychiatry Psychiatr Epidemiol 2000 Nov;35(11):500-507. [doi: 10.1007/s001270050271] [Medline: 11197925]
52. Mgutshini T. Risk factors for psychiatric re-hospitalization: an exploration. Int J Ment Health Nurs 2010 Aug;19(4):257-267. [doi: 10.1111/j.1447-0349.2009.00664.x] [Medline: 20618526]
53. DeNiro DA. Perceived alienation in individuals with residual-type schizophrenia. Issues Ment Health Nurs 1995;16(3):185-200. [doi: 10.3109/01612849509006934] [Medline: 7759231]
54. Arnold C, Villagonzalo KA, Meyer D, Farhall J, Foley F, Kyrios M, et al. Predicting engagement with an online psychosocial intervention for psychosis: exploring individual- and intervention-level predictors. Internet Interv 2019 Aug 15;18:100266 [FREE Full text] [doi: 10.1016/j.invent.2019.100266] [Medline: 31890619]
55. Inkster B, Stillwell D, Kosinski M, Jones P. A decade into Facebook: where is psychiatry in the digital age? Lancet Psychiatry 2016 Nov;3(11):1087-1090. [doi: 10.1016/S2215-0366(16)30041-4] [Medline: 27794373]
56. Cole-Lewis H, Kershaw T. Text messaging as a tool for behavior change in disease prevention and management. Epidemiol Rev 2010;32(1):56-69 [FREE Full text] [doi: 10.1093/epirev/mxq004] [Medline: 20354039]
57. Woolford SJ, Barr KL, Derry HA, Jepson CM, Clark SJ, Strecher VJ, et al. OMG do not say LOL: obese adolescents' perspectives on the content of text messages to enhance weight loss efforts. Obesity (Silver Spring) 2011 Dec;19(12):2382-2387 [FREE Full text] [doi: 10.1038/oby.2011.266] [Medline: 21869762]
58. Napolitano MA, Hayes S, Bennett GG, Ives AK, Foster GD. Using Facebook and text messaging to deliver a weight loss program to college students. Obesity (Silver Spring) 2013 Jan;21(1):25-31 [FREE Full text] [doi: 10.1002/oby.20232] [Medline: 23505165]
59. Schrank B, Sibitz I, Unger A, Amering M. How patients with schizophrenia use the internet: qualitative study. J Med Internet Res 2010 Dec 19;12(5):e70 [FREE Full text] [doi: 10.2196/jmir.1550] [Medline: 21169176]
60. Ainsworth J, Palmier-Claus JE, Machin M, Barrowclough C, Dunn G, Rogers A, et al. A comparison of two delivery modalities of a mobile phone-based assessment for serious mental illness: native smartphone application vs text-messaging only implementations. J Med Internet Res 2013 Apr 05;15(4):e60 [FREE Full text] [doi: 10.2196/jmir.2328] [Medline: 23563184]
61. Firth J, Torous J. Smartphone apps for schizophrenia: a systematic review. JMIR Mhealth Uhealth 2015 Nov 06;3(4):e102 [FREE Full text] [doi: 10.2196/mhealth.4930] [Medline: 26546039]
62. Gay K, Torous J, Joseph A, Pandya A, Duckworth K. Digital technology use among individuals with schizophrenia: results of an online survey. JMIR Ment Health 2016 May 04;3(2):e15 [FREE Full text] [doi: 10.2196/mental.5379] [Medline: 27146094]
63. Sharp IR, Kobak KA, Osman DA. The use of videoconferencing with patients with psychosis: a review of the literature. Ann Gen Psychiatry 2011 Apr 18;10(1):14 [FREE Full text] [doi: 10.1186/1744-859X-10-14] [Medline: 21501496]
64. Depp CA, Mausbach B, Granholm E, Cardenas V, Ben-Zeev D, Patterson TL, et al. Mobile interventions for severe mental illness: design and preliminary data from three approaches. J Nerv Ment Dis 2010 Oct;198(10):715-721 [FREE Full text] [doi: 10.1097/NMD.0b013e3181f49ea3] [Medline: 20921861]
65. Kimhy D, Corcoran C. Use of Palm computer as an adjunct to cognitive-behavioural therapy with an ultra-high-risk patient: a case report. Early Interv Psychiatry 2008 Nov;2(4):234-241 [FREE Full text] [doi: 10.1111/j.1751-7893.2008.00083.x] [Medline: 19884956]
66. Granholm E, Ben-Zeev D, Link PC, Bradshaw KR, Holden JL. Mobile Assessment and Treatment for Schizophrenia (MATS): a pilot trial of an interactive text-messaging intervention for medication adherence, socialization, and auditory hallucinations. Schizophr Bull 2012 May;38(3):414-425 [FREE Full text] [doi: 10.1093/schbul/sbr155] [Medline: 22080492]
67. Hatfield E, Rapson RL. From pen pals to chat rooms: the impact of social media on Middle Eastern Society. SpringerPlus 2015 Jun 12;4(1):254. [doi: 10.1186/s40064-015-1033-4] [Medline: 26090303]
68. Oinas-Kukkonen H, Harjumaa M. Persuasive Systems Design: key issues, process model, and system features. Commun Assoc Inf Syst 2009;24(1):28. [doi: 10.17705/1cais.02428]
69. Ng MM, Firth J, Minen M, Torous J. User engagement in mental health apps: a review of measurement, reporting, and validity. Psychiatr Serv 2019 Jul 01;70(7):538-544 [FREE Full text] [doi: 10.1176/appi.ps.201800519] [Medline: 30914003]
70. Bradstreet S. Developing engaging online interventions for people with psychosis. The Mental Elf. 2019. URL: https://www.nationalelfservice.net/treatment/digital-health/developing-engaging-online-interventions-for-people-with-psychosis/ [accessed 2022-05-16]
71. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the trial: systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. J Med Internet Res 2018 Jun 06;20(6):e199 [FREE Full text] [doi: 10.2196/jmir.9275] [Medline: 29875089]

72. Kuo KM, Chen YC, Talley PC, Huang CH. Continuance compliance of privacy policy of electronic medical records: the roles of both motivation and habit. BMC Med Inform Decis Mak 2018 Dec 18;18(1):135 [FREE Full text] [doi: 10.1186/s12911-018-0722-7] [Medline: 30563500]
73. Kanter JW, Manos RC, Bowe WM, Baruch DE, Busch AM, Rusch LC. What is behavioral activation? A review of the empirical literature. Clin Psychol Rev 2010 Aug;30(6):608-620. [doi: 10.1016/j.cpr.2010.04.001] [Medline: 20677369]
74. Granovetter MS. The strength of weak ties. Am J Sociol 1973 May;78(6):1360-1380. [doi: 10.1086/225469]
75. Morgenroth T, Ryan MK, Peters K. The motivational theory of role modeling: how role models influence role aspirants' goals. Rev Gen Psychol 2015 Dec 01;19(4):465-483. [doi: 10.1037/gpr0000059]
76. Deci EL, Ryan RM. The "What" and "Why" of goal pursuits: human needs and the self-determination of behavior. Psychol Inq 2000 Oct;11(4):227-268. [doi: 10.1207/s15327965pli1104_01]
77. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]
78. Flay BR, Petraitis J. The theory of triadic influence: a new theory of health behavior with implications for preventive interventions. Adv Med Sociol 1994;4:19-44. [doi: 10.4135/9781412952576.n210]
79. Bassett Jr DR, Toth LP, LaMunion SR, Crouter SE. Step counting: a review of measurement considerations and health-related applications. Sports Med 2017 Jul;47(7):1303-1315 [FREE Full text] [doi: 10.1007/s40279-016-0663-1] [Medline: 28005190]
80. Tison GH, Marcus GM. Will the smartphone become a useful tool to promote physical activity? Lancet Digit Health 2019 Nov;1(7):e322-e323 [FREE Full text] [doi: 10.1016/S2589-7500(19)30154-2] [Medline: 33323203]
81. Evans WD, Harrington C, Patchen L, Andrews V, Gaminian A, Ellis LP, et al. Design of a novel digital intervention to promote healthy weight management among postpartum African American women. Contemp Clin Trials Commun 2019 Oct 3;16:100460 [FREE Full text] [doi: 10.1016/j.conctc.2019.100460] [Medline: 31650078]
82. Lindsay S, Cagliostro E, Leck J, Stinson J. A 12-week electronic mentoring employment preparation intervention for youth with physical disabilities: pilot feasibility randomized controlled trial. JMIR Pediatr Parent 2019 Mar 29;2(1):e12088 [FREE Full text] [doi: 10.2196/12088] [Medline: 31518315]
83. McGregor KM, Crosson B, Mammino K, Omar J, García PS, Nocera JR. Influences of 12-week physical activity interventions on TMS measures of cortical network inhibition and upper extremity motor performance in older adults-a feasibility study. Front Aging Neurosci 2017 Jan 4;9:422 [FREE Full text] [doi: 10.3389/fnagi.2017.00422] [Medline: 29354049]
84. Peyman N, Rezai-Rad M, Tehrani H, Gholian-Aval M, Vahedian-Shahroodi M, Heidarian Miri H. Digital Media-based Health Intervention on the promotion of Women's physical activity: a quasi-experimental study. BMC Public Health 2018 Jan 15;18(1):134 [FREE Full text] [doi: 10.1186/s12889-018-5025-5] [Medline: 29334970]
85. Shvedko AV, Thompson JL, Greig CA, Whittaker AC. Physical Activity Intervention for Loneliness (PAIL) in community-dwelling older adults: protocol for a feasibility study. Pilot Feasibility Stud 2018 Dec 19;4:187 [FREE Full text] [doi: 10.1186/s40814-018-0379-0] [Medline: 30598833]
86. Tsai HH, Tsai YF. Changes in depressive symptoms, social support, and loneliness over 1 year after a minimum 3-month videoconference program for older nursing home residents. J Med Internet Res 2011 Nov 15;13(4):e93 [FREE Full text] [doi: 10.2196/jmir.1678] [Medline: 22086660]
87. Lal S, Dell'Elce J, Tucci N, Fuhrer R, Tamblyn R, Malla A. Preferences of young adults with first-episode psychosis for receiving specialized mental health services using technology: a survey study. JMIR Ment Health 2015 May 20;2(2):e18 [FREE Full text] [doi: 10.2196/mental.4400] [Medline: 26543922]
88. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013 Apr 18;15(4):e86 [FREE Full text] [doi: 10.2196/jmir.2583] [Medline: 23598614]
89. Schwartz B. Self-determination. The tyranny of freedom. Am Psychol 2000 Jan;55(1):79-88. [doi: 10.1037//0003-066x.55.1.79] [Medline: 11392868]
90. Crutzen R, Cyr D, de Vries NK. The role of user control in adherence to and knowledge gained from a website: randomized comparison between a tunneled version and a freedom-of-choice version. J Med Internet Res 2012 Mar 09;14(2):e45 [FREE Full text] [doi: 10.2196/jmir.1922] [Medline: 22532074]
91. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to Web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
92. Priebe S, Pavlickova H, Eldridge S, Golden E, McCrone P, Ockenden N, et al. Effectiveness of one-to-one volunteer support for patients with psychosis: protocol of a randomised controlled trial. BMJ Open 2016 Aug 03;6(8):e011582 [FREE Full text] [doi: 10.1136/bmjopen-2016-011582] [Medline: 27489153]
93. Armstrong J. How effective are minimally trained/experienced volunteer mental health counsellors? Evaluation of CORE outcome data. Couns Psychother Res 2010 Mar;10(1):22-31. [doi: 10.1080/14733140903163284]
94. Kingdon DG, Turkington D, Collis J, Judd M. Befriending: cost-effective community care. Psychiatr Bull 1989;13(7):350-351. [doi: 10.1192/pb.13.7.350]
95. McGowan B, Jowett C. Promoting positive mental health through befriending. Int J Ment Health Promot 2003;5(2):12-24. [doi: 10.1080/14623730.2003.9721902]

96. Tombs D, Stowers C, Faibank S, Akrill T. A befriending service for individuals with complex, enduring mental health problems. Clin Psychol 2003 Aug;BPS Montlhy Division(28):33-36.
97. Brackhane R, Strehl C, Wurzer I. [Lay help in rehabilitation of psychiatrically handicapped patients--report of two comparative empirical studies]. Rehabilitation (Stuttg) 1990 Nov;29(4):254-260. [Medline: 2291029]
98. Rotondi AJ, Sinkule J, Haas GL, Spring MB, Litschge CM, Newhill CE, et al. Designing websites for persons with cognitive deficits: design and usability of a psychoeducational intervention for persons with severe mental illness. Psychol Serv 2007 Aug;4(3):202-224 [FREE Full text] [doi: 10.1037/1541-1559.4.3.202] [Medline: 26321884]
99. Sundram F, Corattur T, Dong C, Zhong K. Motivations, expectations and experiences in being a mental health helplines volunteer. Int J Environ Res Public Health 2018 Sep 27;15(10):2123 [FREE Full text] [doi: 10.3390/ijerph15102123] [Medline: 30261682]
100. Margaret ME, Goldberg MS, Anne S, Evans SM, Katherine H, Cole SM. The utilization and training of volunteers in a psychiatric setting. Br J Soc Work 1973:55-63. [doi: 10.1093/oxfordjournals.bjsw.a056392]
101. Lobban F, Dodd AL, Sawczuk AP, Asar O, Dagnan D, Diggle PJ, et al. Assessing feasibility and acceptability of Web-based enhanced relapse prevention for bipolar disorder (ERPonline): a randomized controlled trial. J Med Internet Res 2017 Mar 24;19(3):e85 [FREE Full text] [doi: 10.2196/jmir.7008] [Medline: 28341619]
102. Fry JP, Neff RA. Periodic prompts and reminders in health promotion and health behavior interventions: systematic review. J Med Internet Res 2009 May 14;11(2):e16 [FREE Full text] [doi: 10.2196/jmir.1138] [Medline: 19632970]
103. Whitton AE, Proudfoot J, Clarke J, Birch MR, Parker G, Manicavasagar V, et al. Breaking open the black box: isolating the most potent features of a Web and mobile phone-based intervention for depression, anxiety, and stress. JMIR Ment Health 2015 Mar 4;2(1):e3 [FREE Full text] [doi: 10.2196/mental.3573] [Medline: 26543909]
104. Alkhaldi G, Hamilton FL, Lau R, Webster R, Michie S, Murray E. The effectiveness of prompts to promote engagement with digital interventions: a systematic review. J Med Internet Res 2016 Jan 08;18(1):e6 [FREE Full text] [doi: 10.2196/jmir.4790] [Medline: 26747176]
105. Hsu A, Blandford A. Designing for psychological change: individuals' reward and cost valuations in weight management. J Med Internet Res 2014 Jun 26;16(6):e138 [FREE Full text] [doi: 10.2196/jmir.3009] [Medline: 24972304]
106. Walsh S, Golden E, Priebe S. Systematic review of patients' participation in and experiences of technology-based monitoring of mental health symptoms in the community. BMJ Open 2016 Jun 21;6(6):e008362 [FREE Full text] [doi: 10.1136/bmjopen-2015-008362] [Medline: 27329437]
107. Maj M. Physical health care in persons with severe mental illness: a public health and ethical priority. World Psychiatry 2009 Feb;8(1):1-2 [FREE Full text] [doi: 10.1002/j.2051-5545.2009.tb00196.x] [Medline: 19293947]
108. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003 Aug;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]
109. Ling T. Evaluating complex and unfolding interventions in real time. Evaluation 2012 Jan 22;18(1):79-91. [doi: 10.1177/1356389011429629]
110. Fletcher A, Jamal F, Moore G, Evans RE, Murphy S, Bonell C. Realist complex intervention science: applying realist principles across all phases of the Medical Research Council framework for developing and evaluating complex interventions. Evaluation (Lond) 2016 Jul;22(3):286-303 [FREE Full text] [doi: 10.1177/1356389016652743] [Medline: 27478401]
111. Biagianti B, Hidalgo-Mazzei D, Meyer N. Developing digital interventions for people living with serious mental illness: perspectives from three mHealth studies. Evid Based Ment Health 2017 Nov;20(4):98-101 [FREE Full text] [doi: 10.1136/eb-2017-102765] [Medline: 29025862]
112. Bradbury K, Steele M, Corbett T, Geraghty AW, Krusche A, Heber E, et al. Developing a digital intervention for cancer survivors: an evidence-, theory- and person-based approach. NPJ Digit Med 2019 Sep 2;2:85 [FREE Full text] [doi: 10.1038/s41746-019-0163-4] [Medline: 31508496]
113. Dack C, Ross J, Stevenson F, Pal K, Gubert E, Michie S, et al. A digital self-management intervention for adults with type 2 diabetes: combining theory, data and participatory design to develop HeLP-Diabetes. Internet Interv 2019 Sep;17:100241 [FREE Full text] [doi: 10.1016/j.invent.2019.100241] [Medline: 31372349]
114. Kassavou A, Houghton V, Edwards S, Brimicombe J, Sutton S. Development and piloting of a highly tailored digital intervention to support adherence to antihypertensive medications as an adjunct to primary care consultations. BMJ Open 2019 Jan 06;9(1):e024121 [FREE Full text] [doi: 10.1136/bmjopen-2018-024121] [Medline: 30613027]
115. Kujala S. User involvement: a review of the benefits and challenges. Behav Inf Technol 2003 Jan;22(1):1-16. [doi: 10.1080/01449290301782]
116. De Vito Dabbs A, Myers BA, Mc Curry KR, Dunbar-Jacob J, Hawkins RP, Begey A, et al. User-centered design and interactive health technologies for patients. Comput Inform Nurs 2009;27(3):175-183 [FREE Full text] [doi: 10.1097/NCN.0b013e31819f7c7c] [Medline: 19411947]
117. Hoddinott P. A new era for intervention development studies. Pilot Feasibility Stud 2015 Oct 26;1:36 [FREE Full text] [doi: 10.1186/s40814-015-0032-0] [Medline: 27965814]
118. Kok G, Schaalma H. Using theory in psychological interventions. In: Michie S, Abraham C, editors. Health Psychology in Practice. Hoboken, NJ, USA: Blackwell Publishing; 2004:203-229.

119. Lewin S, Hendry M, Chandler J, Oxman AD, Michie S, Shepperd S, et al. Assessing the complexity of interventions within systematic reviews: development, content and use of a new tool (iCAT_SR). BMC Med Res Methodol 2017 Apr 26;17(1):76 [FREE Full text] [doi: 10.1186/s12874-017-0349-x] [Medline: 28446138]
120. Logic Model Volunteer project support 2015. Wikimedia. 2015. URL: https://upload.wikimedia.org/wikipedia/commons/b/b8/Logic_Model_Volunteer_project_support_2015_final.pdf [accessed 2022-05-16]
121. Volunteer Management: Using a Logic Model - The National Long-Term Care Ombudsman Resource Center. National Association of Area Agencies on Aging. 2017. URL: https://ltcombudsman.org/uploads/files/support/vol-mgt-oct-24.pdf [accessed 2022-05-16]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
MRC: Medical Research Council
REC: Research Ethics Committee
SMI: severe mental illness

Edited by A Mavragani; submitted 20.11.21; peer-reviewed by J Ibrahim, M Brown; comments to author 25.01.22; revised version received 16.03.22; accepted 17.03.22; published 02.06.22.

Please cite as:
Pinto da Costa M, Phone Pal Advisory Groups
An Intervention to Connect Patients With Psychosis and Volunteers via Smartphone (the Phone Pal): Development Study
JMIR Form Res 2022;6(6):e35086
URL: https://formative.jmir.org/2022/6/e35086
doi:10.2196/35086
PMID:35653171

Original Paper

A Yoga Exercise App Designed for Patients With Axial Spondylarthritis: Development and User Experience Study

Minh Tam Truong[1]; Obioma Bertrand Nwosu[1]; Maria Elena Gaytan Torres[1]; Maria Paula Segura Vargas[1]; Ann-Kristin Seifer[1], MSc; Marlies Nitschke[1], PhD; Alzhraa A Ibrahim[2], MSc; Johannes Knitza[3], MD, Dr med; Martin Krusche[4], Dr med; Bjoern M Eskofier[1], Prof Dr; Georg Schett[3], Prof Dr; Harriet Morf[3], Dr med

[1]Machine Learning and Data Analytics Lab, Faculty of Engineering, Department of Artificial Intelligence in Biomedical Engineering, Friedrich-Alexander University Erlangen-Nürnberg, Erlangen, Germany

[2]Computer Science Department, Faculty of Computers and Information, Assiut University, Assiut, Egypt

[3]Department of Internal Medicine 3, Rheumatology and Immunology, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nürnberg, Erlangen, Germany

[4]Rheumatology and Clinical Immunology, Charité Universitätsmedizin Berlin, Berlin, Germany

Corresponding Author:
Harriet Morf, Dr med
Department of Internal Medicine 3, Rheumatology and Immunology
University Hospital Erlangen
Friedrich-Alexander University Erlangen-Nürnberg
Ulmenweg 18
Erlangen, 91054
Germany
Phone: 49 9131 8543023
Email: harriet.morf@uk-erlangen.de

Abstract

Background: Besides anti-inflammatory medication, physical exercise represents a cornerstone of modern treatment for patients with axial spondyloarthritis (AS). Digital health apps (DHAs) such as the yoga app *YogiTherapy* could remotely empower patients to autonomously and correctly perform exercises.

Objective: This study aimed to design and develop a smartphone-based app, *YogiTherapy*, for patients with AS. To gain additional insights into the usability of the graphical user interface (GUI) for further development of the app, this study focused exclusively on evaluating users' interaction with the GUI.

Methods: The development of the app and the user experience study took place between October 2020 and March 2021. The DHA was designed by engineering students, rheumatologists, and patients with AS. After the initial development process, a pilot version of the app was evaluated by 5 patients and 5 rheumatologists. The participants had to interact with the app's GUI and complete 5 navigation tasks within the app. Subsequently, the completion rate and experience questionnaire (attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty) were completed by the patients.

Results: The results of the posttest questionnaires showed that most patients were already familiar with digital apps (4/5, 80%). The task completion rates of the usability test were 100% (5/5) for the tasks T1 and T2, which included selecting and starting a yoga lesson and navigating to an information page. Rheumatologists indicated that they were even more experienced with digital devices (2/5, 40% experts; 3/5, 60% intermediates). In this case, they scored task completion rates of 100% (5/5) for all 5 usability tasks T1 to T5. The mean results from the User Experience Questionnaire range from −3 (most negative) to +3 (most positive). According to rheumatologists' evaluations, attractiveness (mean 2.267, SD 0.401) and stimulation (mean 2.250, SD 0.354) achieved the best mean results compared with dependability (mean 2.000, SD 0.395). Patients rated attractiveness at a mean of 2.167 (SD 0.565) and stimulation at a mean of 1.950 (SD 0.873). The lowest mean score was reported for perspicuity (mean 1.250, SD 1.425).

Conclusions: The newly developed and tested DHA *YogiTherapy* demonstrated moderate usability among rheumatologists and patients with rheumatic diseases. The app can be used by patients with AS as a complementary treatment. The initial evaluation of the GUI identified significant usability problems that need to be addressed before the start of a clinical evaluation. Prospective trials are also needed in the second step to prove the clinical benefits of the app.

(JMIR Form Res 2022;6(6):e34566) doi:10.2196/34566

KEYWORDS

spondylarthritis; digital health application; yoga; usability; patient empowerment; mobile health; mHealth; health applications; smartphone; physical exercise; wellness; mobile phone

Introduction

Background

Axial spondylarthritis (AS) is a systemic rheumatic disease that causes chronic inflammation that affects the spinal column and adjacent joints. Affected individuals have inflammatory back pain and decreased spinal mobility because of spinal stiffening. The symptoms are related to pain and decreased function and quality of life. If inflammation progresses, deformities and ossification of the spine may occur. To mitigate the progression of the disease, it is not sufficient to only follow pharmacological and nonpharmacological treatments. Drug options are limited; although with the approval of biologics for the disease, there has been a marked improvement in the treatment therapy. Nevertheless, nonpharmacological therapy, such as physical therapy and self-directed physical exercises, is a complementary part of treatment and can be easily performed by the patients if there is sufficient motivation for physical exercise on the part of the patients [1,2]. Therefore, it is essential to foster active patient participation through self-management interventions, which give patients the ability to actively manage their symptoms and treatment. Active engagement has a positive impact on clinical patient outcome [3]. Digital health apps (DHAs) focusing on exercise can be one way to easily motivate patients in their daily lives [4]. In a survey on the usefulness and willingness to use and pay for a rheumatic self-management app, many patients had a positive attitude toward it. However, the success of the interventions in real life is limited by poor patient-physician coordination and reduced access to self-management programs for patients [5]. These circumstances have particularly worsened since the outbreak of the COVID-19 pandemic.

Objectives

Health insurance companies in Germany offer various preventive services such as web-based coaching and physical exercise courses [6]. The courses provide a variety of services, ranging from stress management to back and endurance training, and take place at a specific location. Web-based coaching programs are web-based courses and fitness apps that help patients track their diet, improve their physical condition, and provide exercises and information resources on specific topics, such as their back health [7,8]. These services are widely accepted by patients. However, they generally target patients with various musculoskeletal conditions who need support to overcome the challenges of physical inactivity, sedentary behavior, and unhealthy diets. As these offerings tend to be aimed at a broad community, they often overlook the needs of patients with AS or specific information that is important to them. There is a lack of solutions that are aimed exclusively at patients with AS and meet the demands mentioned earlier [9]. The possible solutions are presented below.

Mobile devices, such as smartphones and tablet computers, are widely used by patients with rheumatic diseases and can assist the user in daily living without special training. Mobile health apps installed on mobile devices offer great potential to improve patients' self-management capabilities [5,10]. No therapeutic DHA for patients with rheumatic diseases could be identified in a recent analysis, although patients clearly stated this need [11].

In a survey about the use [12], preferences, and perceptions of DHAs in the era of COVID-19, the attitudes of patients and rheumatologists changed positively toward the implementation of DHAs (38%) in clinical care. DHA use had increased because of COVID-19 (29%). Most patients (74%) and rheumatologists (76%) believed that DHAs are useful in the management of rheumatic and musculoskeletal diseases and felt confident in using these DHAs (90% and 86%, respectively) [12]. In light of this, it is a problem that currently there are no approved DHAs for rheumatological diseases except for individual apps [11,13].

For instance, one existing app on the market calculates the level of AS in a given patient to measure the severity of the disease [14]. This app additionally provides information on the classification, diagnosis, progression, and treatment of the disease [15]. In addition, there are web solutions that focus on displaying exercises suitable for patients with AS [16] or providing educational information about the disease [17]. Another interesting solution helps people with rheumatoid arthritis to monitor their condition by tracking personal facts (eg, inability to work, morning stiffness, and times of infections) [18]. If we focus on DHAs that mainly cover the fitness sector or are specifically designed for AS, we find 3 competitors. The first is the app Assessment of SpondyloArthritis (ASAS) [19]. In this paper, patients can calculate the Ankylosing Spondylitis Disease Activity Score and Bath Ankylosing Spondylitis Disease Activity Index (BASDAI). In addition, there are informative parts about the disease and questionnaires. The second competitor is the fitness app Gymondo [20]. It focuses on fitness and nutrition. Exercises lasting approximately 50 minutes are designed to prevent back pain. A fitness test can be performed as a bonus. However, there is a lack of special exercises for patients with rheumatic diseases. Another app is Kaia Health, which is specifically aimed at back pain [21] and focuses on pain therapy. Participants receive feedback via pose estimation.

Notwithstanding these, a solution specifically for patients with AS, ideally combining the benefits of fitness exercise, education, and progress tracking, is not yet available.

Nonpharmacological therapy for AS should focus on muscle building and stretching. Yoga combines these 2 aspects and can be adapted for different levels of difficulty. Some studies on patients with rheumatic diseases have shown improvements in function and disease activity. Experience of pain and sleep behavior can also be positively influenced by yoga [22,23].

Moreover, there already exist individual studies investigating the influence of yoga on patients with AS (ClinicalTrials.gov identifier: NCT04281238; DRKS-ID: DRKS00025215 [24]).

This motivated us to develop the app *YogiTherapy* to improve current rheumatology care. The aim of the app is to promote flexible exercising in a home-based environment and to improve the quality of coordination of care between patients and physicians through self-monitoring.

This study is the first step in DHA development and mainly focuses on the technical aspects and the evaluation of the graphical user interface (GUI).

To ensure that the developed app is appropriately tailored to patients' needs and thus can later be successfully implemented as digital health care technology, it is important to already evaluate its user experience in an early development stage. Schrepp et al [25] define user experience as a holistic concept that combines classical usability criteria with hedonic quality criteria. Therefore, this pilot study aimed to identify usability issues of *YogiTherapy*'s GUI and to evaluate its user experience using the User Experience Questionnaire (UEQ). Clinical feasibility and evaluation were not examined in this study.

Methods

Overview

This section highlights the development process and provides an overview of the basic structure of the apps under investigation. Furthermore, it presents the methods to evaluate the user experience as well as digital habits and use of smartphone-based apps.

Development of the App YogiTherapy

YogiTherapy was implemented as a DHA by engineering students from the Machine Learning and Data Analytics Lab at the University of Erlangen-Nuremberg (FAU) and physicians at the University Hospital in Erlangen. Feedback from patients with rheumatic diseases, participating in live yoga classes, was also considered during the conceptional design process. This enabled us to adapt all yoga positions to patients' movement limitations.

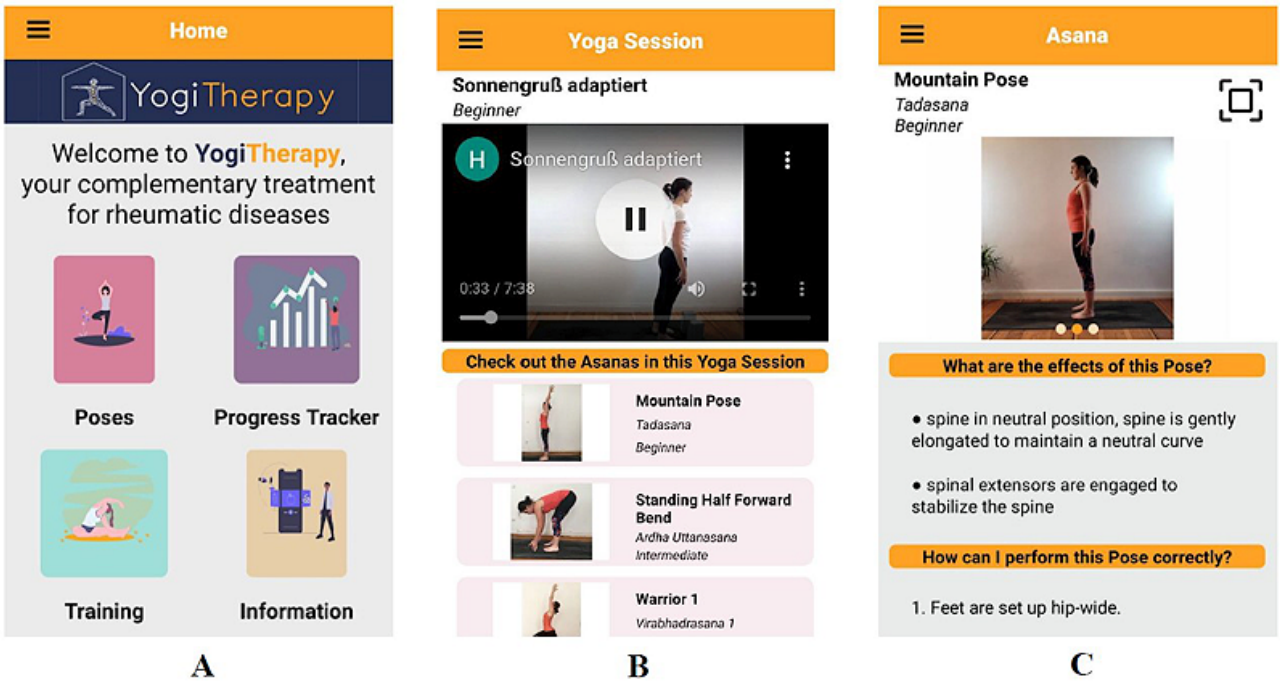
We wanted to create a cross-platform mobile app to target users of the most popular and widely available mobile platforms, iOS and Android [26]. Thus, we chose the open-source cross-platform mobile app framework *React Native*.

We structured the app into 5 sections: training, pose estimation, assessment tests, progress tracker, and information.

The home screen of the app provides easy access to the poses, training, progress, and information sections (Figure 1A). The following paragraphs present the functionalities of the app.

The purpose of the training section of the app was to provide users with different exercise resources targeted specifically for patients with AS. Compared with other available exercising apps, the yoga poses and instructions were provided by a certified yoga instructor with a professional background in rheumatology to ensure that they are adequately adapted to patients' needs.

Figure 1. Home screen (A) and training section (B and C) in the app. (B) Video tutorial and its corresponding static yoga poses. (C) Instruction page of a selected yoga pose.

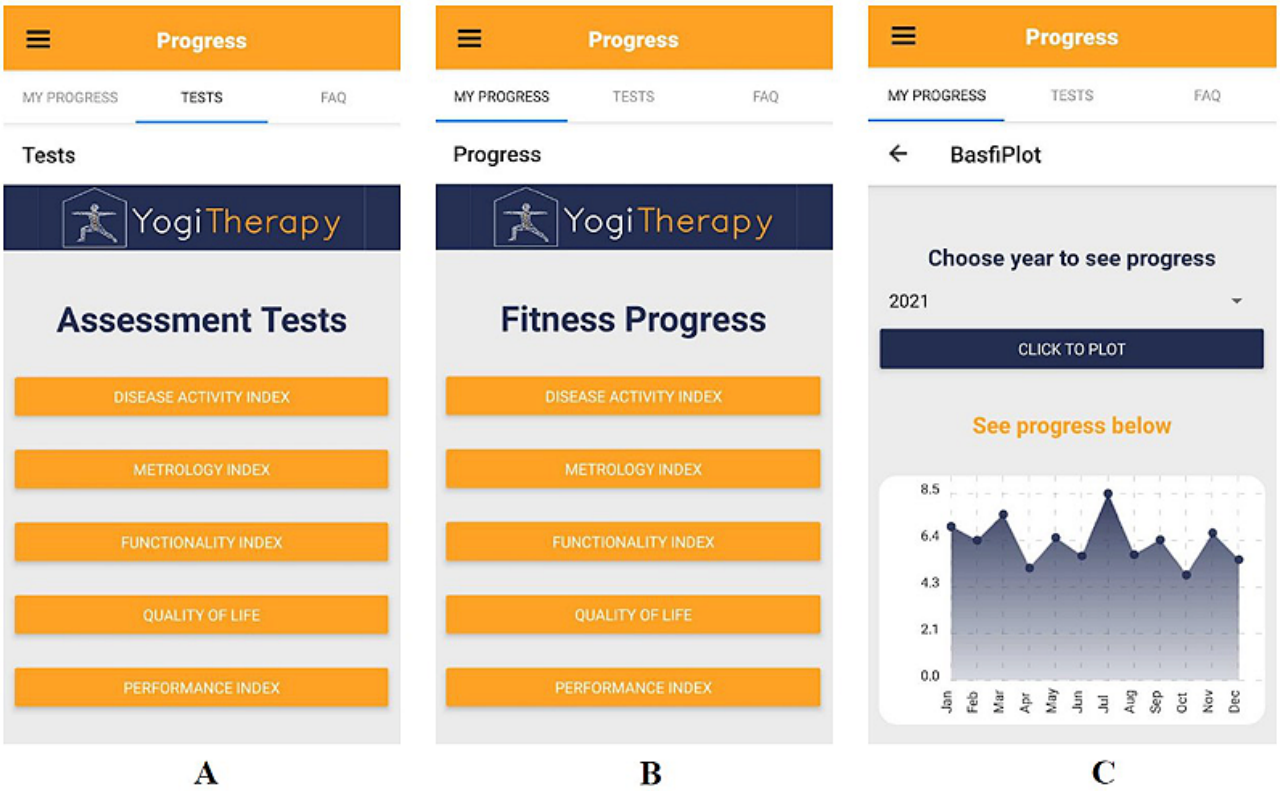


The exercise materials were shared in the form of video tutorials and images. The videos, which can be played directly inside the app (Figure 1B), were divided into beginner and intermediate levels, depending on movement restrictions and fitness level. In addition, the training section offered detailed instructions and visual depictions of each pose (Figure 1C). This allowed

the users to learn and perform each pose in advance without sustaining injuries. With the introduction of pose estimation (computer vision techniques that estimate the spatial location of joints from image or video data), we were able to provide users with real-time feedback during the performance of single static yoga poses. The feedback pipeline consisted of 2 steps: pose estimation and pose assessment. The pose estimation relied on a pretrained open-source PoseNet model (version 2.2.1; MobileNet version 2.0.4) that was fed with a live stream of the camera view to detect 17 key points (eg, left ankle) and output spatial information. This spatial information of the detected key points was then compared with the extracted key points of a reference image of the yoga pose using cosine similarity and Euclidean distance. As soon as the cosine similarity between both key point data sets exceeds a defined threshold, the detected skeleton in the camera view turns green to indicate good accordance with the reference pose image. In the case of low similarity, a red-colored skeleton signals the user to keep adjusting the pose.

We incorporated a progress tracker to give users the possibility to self-monitor their symptoms and disease status and track their own progress over time (Figure 2). This allows users to take different assessment tests (a detailed explanation of the chosen assessment tests is provided in Multimedia Appendix 1 [27-29]) and then review their health progress over time.

Figure 2. Progress tracker section in the app. (A) Overview screen of assessment tests. (B) Overview screen for test results. (C) Visual depiction of the annual test result of the Bath Ankylosing Spondylitis Functional Index test.



We included 3 subjective test instruments, the Bath Ankylosing Spondylitis Functional Index (BASFI), the BASDAI, and the Ankylosing Spondylitis quality of life, and 2 objective test instruments, the Bath Ankylosing Spondylitis Metrology Index (BASMI) and the Ankylosing Spondylitis Performance Index.

Exactly how often these tests need to be administered to collect sufficient reliable data over a long treatment period will be the subject of a prospective clinical evaluation. The test results were stored locally on a mobile device to comply with high data privacy standards. This personal medical record can be shared with consulting rheumatologists at follow-up appointments if desired. Patients can access and view their entire medical history on a progress tracker dashboard. The dashboard was structured according to the type of test (Figure 2B), so that it was possible to display the monthly means for each test in the form of an annual line or bar chart (Figure 2C).

For this study, we only worked with the Android version of the system and delivered it in the package file format Android Package (APK).

Ethics Approval

All participants provided consent, and the study was approved by the ethics committee of the Medical Faculty of the University of Erlangen-Nuremberg, Germany (Antrag 8_ 21 B).

Study Design

For this pilot user experience evaluation study, we used a mixed methods design and collected qualitative and quantitative data. Participants performed a formative usability test immediately followed by a posttest questionnaire to quantify the user

experience of *YogiTherapy*'s GUI. This pilot study was conducted in March 2021, and 10 subjects were recruited within 2 weeks through the Department of Rheumatology and Immunology at the University Hospital Erlangen. Participation was voluntary, and there was no preselection in the form of screening. To be included in the study, participants had to be diagnosed with a rheumatic disease or be a rheumatologist. We chose to not limit patient recruitment to only patients with AS because this study did not aim to investigate the clinical feasibility or benefits of the app as a DHA for patients with AS but to evaluate the GUI during the continuing development process. Moreover, we aimed to evaluate the app design from all target users' perspectives. This includes not only patients but also rheumatologists, who have a special mediating role and can take advantage of the continuous health data provided by the app to adjust the treatment. Therefore, we deliberately recruited 5 patients and 5 rheumatologists to explore the differences in the measured user experience between the 2 main target user groups. The general exclusion criterion was lack of basic English skills, as it is the main language of the first app prototype.

This pilot study was part of the first development cycle and served as a first insight into the app's user experience. Previous research estimates that approximately 7 to 8 participants in a usability study can uncover approximately 80% of the existing usability issues when each individual participant uncovers an average of 20% of the issues. Usability studies in later development iterations require a larger sample size because it is becoming increasingly difficult to detect issues [30].

Owing to the COVID-19 restrictions and time limitations, one half of the study was conducted in person and the other half of the study was conducted remotely. The patients' group participated in this study, which was conducted at the outpatient department of the Rheumatology and Immunology of the University Hospital Erlangen, whereas the rheumatologists' group exclusively followed the remote study procedure.

In the beginning, each subject was informed about the purpose and process of the study and was asked to provide written consent. The participants were provided with a test smartphone (Google Pixel 3a, Google Inc) on which the app *YogiTherapy* was preinstalled. In the case of the remote study, the procedure participants could either borrow the test smartphone or install the app on their private Android device.

The study procedure was divided into 2 phases. In phase 1, patients with a rheumatological disease and a rheumatologist were given some time (approximately 3 minutes) to explore the app of *YogiTherapy* and its specifications. After exploration, the patients were instructed to complete 5 navigation task scenarios in the app, which were derived from a prospective user context. The task scenarios were as follows:

- T1. Select a yoga video and start the video.
- T2. Go to the information page about nutrition.
- T3. You have been given a Metrology Index (BASMI) of 7.5 by your doctor. Enter the result in the app and save the result.
- T4. Visualize the Metrology Index (BASMI) over time in the form of a graph.
- T5. Take a test to find out your Disease Activity Index (BASDAI) and save your result.

Patients and rheumatologists could skip each task if they experienced difficulties. For each task, the binary outcome (success or fail) was self-reported using a web-based posttest survey. The task completion rate (TCR), which is the percentage of participants who completed the task, was derived from this [22].

Throughout the interaction with the app, the participants were asked to follow the think-aloud protocol and communicate their thoughts, actions, expectations, and observations. The study conductor collected qualitative data by taking field notes on the subject's verbal reflections. The think-aloud protocol for rheumatologists could not be produced within the framework of the remote study.

In posttest phase 2, each test person was administered a web-based UEQ to assess the user experience of the tested product. In addition, the questionnaire contained questions regarding task completion, aspects of the app they liked or disliked, demographics (sex, age, and patient or rheumatologist), and smartphone experience and habits.

The UEQ allows a quick and reliable measurement of the user experience of interactive products. The questionnaire consists of 6 scales with 26 items in total that measure the distinct quality criteria of user experience. It provides answers to the following questions:

- Attractiveness: Do users like or dislike *YogiTherapy*?
- Perspicuity: Is it easy to get familiar with *YogiTherapy*? Is it easy to understand how to use *YogiTherapy*?
- Efficiency: Can users solve their tasks quickly and efficiently? Does the user interface of *YogiTherapy* look organized?
- Dependability: Does the user feel in control of the interaction? Is the interaction with *YogiTherapy* secure and predictable?
- Stimulation: Is it exciting and motivating to use *YogiTherapy*? Does the user feel motivated for the further use of *YogiTherapy*?
- Novelty: Is the design of the product innovative and creative? Does *YogiTherapy* catch the user's attention?

The attractiveness scale captures the user's general impression of the app and is influenced by the remaining 5 scales. The scales perspicuity, efficiency, and dependability provide information about pragmatic (goal-directed) quality aspects, whereas stimulation and novelty describe hedonic (nongoal-directed) quality criteria [25]. The items have the form of a semantic differential, that is, each item is a pair or opposites with a 7-point Likert scale [31,32].

Analysis

The quantitative participant feedback obtained from the UEQ was processed and analyzed using the Excel Data Analysis Tool [33] provided by the authors of the questionnaire. Independent samples 2-tailed *t* tests (assuming unequal variance, α=.05) were used to compare the user group differences for mean values of each scale. To determine the relative strengths and

weaknesses of *YogiTherapy*, the combined UEQ results of both participant and user groups were compared with a benchmark data set containing 246 user experience evaluations of established products [31]. The TCR of the usability test, demographic characteristics, and smartphone use characteristics were examined using a descriptive analysis. Qualitative data obtained from the think-aloud protocol were clustered in terms of the usability problems encountered.

Results

Descriptive Information

In total, 10 participants completed the study procedure and were included in the final analysis. The participant group *Patient* consisted of 5 patients (3 females and 2 males) with a mean age of 59.4 (SD 6.2) years. Furthermore, the patients were diagnosed with axial spondylarthritis, psoriatic arthritis, and rheumatoid arthritis. In the participant group *Rheumatologist*, 5 rheumatologists were recruited with a mean age of 38.8 (SD 10.7) years.

Results of Group Analysis

The TCR values obtained from the formative usability tests for both groups are presented in Table 1. All rheumatologists stated that they had successfully completed all tasks, whereas the patient group particularly struggled with tasks related to the *Progress Tracker* screen in the app. Of the 5 patients, 2 (40%) were unable to complete tasks T3 and T4. Table 2 presents the issues identified using the think-aloud protocol during the patients' interaction with the app. The identified problems were in accordance with the low completion rates for tasks T3 to T5.

Table 1. Task completion rates of both participant groups.

Tasks	Task completion rate, n (%)	
	Patients (n=5)	Rheumatologists (n=5)
T1	5 (100)	5 (100)
T2	5 (100)	5 (100)
T3	3 (60)	5 (100)
T4	3 (60)	5 (100)
T5	4 (80)	5 (100)

Table 2. Usability issues identified by patients' think-aloud protocols.

Aspect	Usability issue
Overall	• German version would be easier to handle • App crashes sometimes • Information pages are too text-heavy • Filter list function should support automatic filtering • Asana list under video was mistaken for an automatic playlist because of design • A smartphone is impractical for exercising, and a tablet would be more appropriate, especially for the videos
Progress tracker	• Top tab navigation (My Progress—Tests-FAQ[a]) was overlooked • Difficulties finding the tab Tests because the default landing page of the Progress Tracker is My Progress and both the tabs look identical (compare Figure 2A and Figure 2B)

[a]*My Progress-Test-FAQ* is the tab where patients can see the disease progress as it was requested in T4. The Metrology Index (BASMI) is visualized over time in the form of a graph.

In the posttest phase, we asked about their experiences and habits using digital apps and smartphones. Overall, 80% (4/5) of the patients considered themselves to be experienced in using these technologies, compared with 20% (1/5) of the beginners. Majority (4/5, 80%) of the patients reported a regular habit of using smartphones and tablets for 4 to 7 days per week. In the rheumatologist group, 100% (5/5) rheumatologists used their smartphone and tablets for 4 to 7 days per week. Moreover, 40% (2/5) of them claimed to be experts in using these technologies, and 60% (3/5) of them self-reported intermediate skills. The participants' feedback was measured using the UEQ, as shown in Figure 3.

Table 3 provides the UEQ results at the 95% CIs. The scale ranges from −3 (most negative evaluation) to +3 (most positive evaluation) [25]. The mean score of both groups was >0.8, indicating a positive evaluation. Rheumatologists rated *YogiTherapy*'s attractiveness, pragmatic, and hedonic qualities equally high, whereas patients found the app very attractive but reported much lower pragmatic and hedonic qualities. The lowest mean score was for the scale perspicuity (mean 1.250, SD 1.425) in the patient group. The most striking observation to emerge from the data comparison was that, on average, the rheumatologists group rated each user experience scale higher than the patient group. The mean score of rheumatologists was 2.0, indicating a very positive evaluation. However, the differences in the means for each UEQ scale between the groups were not statistically significant.

Figure 3. User Experience Questionnaire scale means depending on the participant group. The error bars represent the 95% CIs of the scale mean.

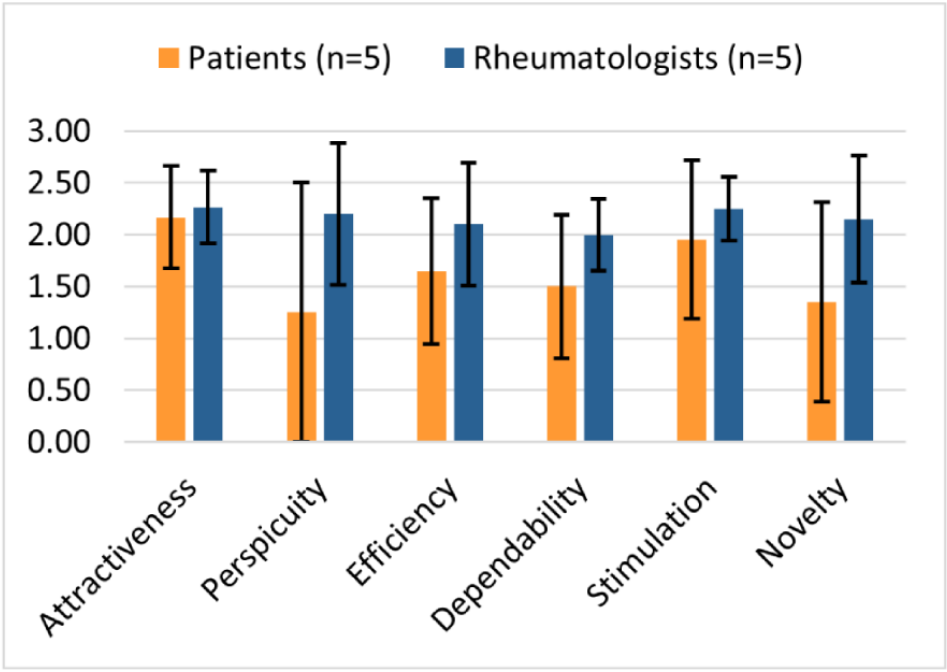
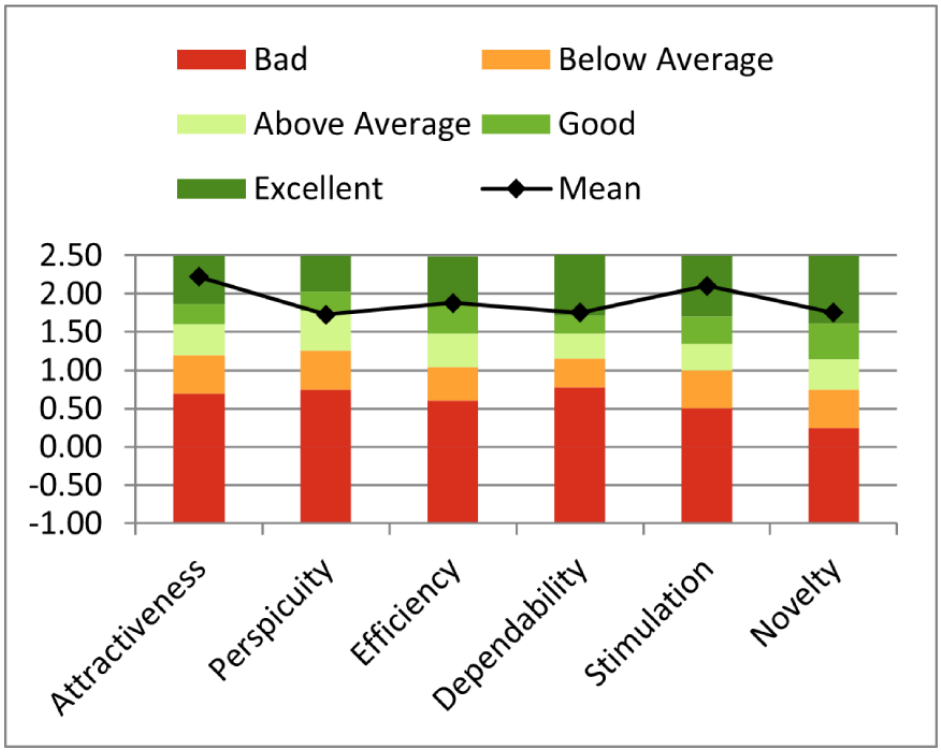


Table 3. User Experience Questionnaire scales of both participant groups: patients and rheumatologists.

Scale	CIs (*P*=.05) per scale					
	Patients (n=5)			Rheumatologists (n=5)		
	Value, mean (SD)	Confidence	95% CI	Value, mean (SD)	Confidence	95% CI
Attractiveness	2.167 (0.565)	0.495	1.671-2.662	2.267 (0.401)	0.352	1.915-2.618
Perspicuity	1.250 (1.425)	1.249	0.001-2.499	2.200 (0.779)	0.682	1.518-2.882
Efficiency	1.650 (0.802)	0.703	0.947-2.353	2.100 (0.675)	0.592	1.508-2.692
Dependability	1.500 (0.791)	0.693	0.807-2.193	2.000 (0.395)	0.346	1.654-2.346
Stimulation	1.950 (0.873)	0.765	1.185-2.715	2.250 (0.354)	0.310	1.940-2.560
Novelty	1.350 (1.098)	0.963	0.387-2.313	2.150 (0.698)	0.612	1.538-2.762

The results of benchmarking are shown in Figure 4. In comparison with established interactive products, *YogiTherapy*'s overall attractiveness, stimulation, novelty, and dependability were classified as excellent. In terms of efficiency, the app achieved the category good. The worst results were achieved for the scale perspicuity (mean 1.725, SD 1.193). Here, *YogiTherapy* was above average compared with the benchmark data set.

Figure 4. *YogiTherapy*'s mean User Experience Questionnaire scores (n=10) compared with the benchmark data set.



Discussion

Principal Findings

Overall, the findings of this pilot study were encouraging. *YogiTherapy*'s general attractiveness and hedonic qualities scored high compared with the products in the benchmark data set, suggesting that the app fulfilled users' general expectations in terms of attractiveness, stimulation, and novelty. Furthermore, the benchmark identified pragmatic qualities as the weaknesses of GUI. Compared with established interactive products, it only achieved the category *Good* and *Above Average* for the scales efficiency and perspicuity. As general user experience has grown over time, new products should reach the category *Good* on all scales [31]. The poor efficiency and perspicuity scores can be attributed to patient evaluation.

Their lower ratings for pragmatic qualities may be related to the assigned usability tasks. The associated TCR indicated that the patient group struggled with all tasks related to the *Progress Tracker* section in the app (Table 1). This is also evident from their think-aloud protocols; the poorer success on the tasks might be explained by inadequate English skills and a confusing user interface in the *Progress Tracker* section. Participants under observation felt very insecure about their English and stated that using the app would be much more pleasant and easier if it were in their native language, German. Furthermore, the 3-part tab navigation structure of the *Progress Tracker* screen and the identical design of the tabs *My Progress* and *Tests* led to confusion. All these points might have influenced the participants' views of the app's perspicuity and efficiency. The better performance of rheumatologists on the usability tasks may be attributed to their lower mean age and familiarity with digital apps, as they reported using mobile devices on a daily basis. This might also have a positive impact on their English proficiency, which, in turn, would enable self-confident navigation in the app. As perspicuity and efficiency, alongside attractiveness and stimulation, play a key role in successful adherence to DHAs, the goal should be to reach the *Excellent* category on all these user experience scales. The benchmark and qualitative data clarified that pragmatic qualities must be improved to achieve this objective. We see the results as concrete starting points for optimizing the app in further development iterations. This could be achieved by offering a multilingual app that supports the user's native language and a more illustrated GUI. Icons easily capture the user's attention, are generally comprehensible, and provide essential information on the first sight. In addition, the user interface elements used for navigation in the *Progress Tracker* section require significant revision, so that the tab navigation is more predominant. The revised app should then be evaluated again in a usability test with a larger sample size and additional task scenarios.

Comparison With Prior Work

As stated in the Introduction section, currently there are 3 competitors in the category of back pain and digital apps for physical activity. In ASAS app, patients can calculate their disease activity; however, exercises that can be performed at home were missing in this app. To date, no clinical trial has evaluated the ASAS app along with physical exercise. Gymondo provides the option to perform exercises for back pain; however, there is a lack of focus on rheumatologic conditions, specifically AS. In addition, Gymondo is a fitness app that has no option for controlling or managing diseases. Therefore, no clinical studies have been conducted using this app. Similar to Gymondo, there is no way to collect and track disease

progression through questionnaires in the Kaia Health app. The exercises are also designed for back pain, but they have no relation to rheumatological diseases. As this app focuses primarily on pain, it has already been evaluated in a number of studies. Similar to *YogiTherapy*, patients were asked for feedback, which is very important for the further development of the app [34]. A clinical trial of 180 participants with nonspecific lower back pain demonstrated that the use of the app over 12 weeks resulted in a significant reduction in pain [35].

Strength and Limitations

This pilot study had several limitations. First, it only included a small sample size, which affected the CIs of the UEQ scale means. Therefore, all results from the UEQ need to be interpreted with caution. Second, because of limitations resulting from the COVID-19 pandemic (time, staff, and restrictions), it was not possible to create the same experimental environment and conditions for both participant groups. Rheumatologists followed the remote procedure and self-managed the usability test on either their private smartphone or the borrowed test device. No direct observation was possible, and it remains unclear whether the participants conducted the navigation tasks seriously. Therefore, there was no think-aloud data for the rheumatologist group. Third, in this study, the conductor took field notes instead of audio recordings for the think-aloud protocol. Therefore, it was not possible to perform a comprehensive analysis of the transcribed data. Fourth, this pilot user experience study with a clear focus on the technical aspects of the app included patients with various rheumatic diseases

as participants instead of exclusively recruiting patients with AS for the patient group. For use in clinical trials, the app must be directed to patients with AS. Finally, the patients were not native English speakers, which could have affected their ratings. The German version is currently being developed.

Conclusions and Future Directions

The newly developed and evaluated DHA *YogiTherapy* can be used as a complementary treatment for patients with rheumatological diseases, particularly AS. In contrast to digital apps targeting back pain or physical activity, the designed app is specifically tailored to the needs of rheumatic patients, as it combines physical yoga exercises to improve mobility and continuous self-monitoring of disease progression. The results of the user experience evaluation suggested that *YogiTherapy*'s GUI convinced patients as well as rheumatologists with its innovation, attractiveness, novelty, and dependability. However, it also revealed major usability issues related to the efficiency and perspicuity of the user interface. These will need special attention during revision and reevaluation. Prospective longitudinal clinical trials with a large cohort of patients with AS are needed to assess the actual benefits, safety, and acceptance of the app as well as to evaluate the remote delivery mode of physical exercises compared with face-to-face courses.

Acknowledgments

BME gratefully acknowledges the support of the German Research Foundation (DFG) within the framework of the Heisenberg Professorship Program (grant ES 434/8-1). The collaboration was initiated within the framework of the collaborative research center EmpkinS (DFG CRC 1483). AAI was funded by a full scholarship from the Ministry of Higher Education of the Arab Republic of Egypt. This research was supported by the University of Erlangen-Nuremberg (FAU), University Hospital, and the *YogiTherapy* Team. The authors are thankful to their colleagues (Matthias Zürl, Soukaina Boughanem, and Can Sinsoysal) who provided expertise that greatly assisted the research. The authors would like to specially thank Alexander Beyer for the final revision.

Data Availability

Raw data will be uploaded and freely accessible.

Authors' Contributions

MT, ON, MG, MV, AKS, MN, AAI, and HM conceived the study. MT, ON, MG, and MV developed the theory and performed the computations. MT, AKS, MN, AAI, and HM verified the analytical methods. HM, AKS, MN, and AAI encouraged MT to investigate the usability of the app and supervised the study findings. MT, ON, MG, MV, JK, MK, BE, GS, AKS, MN, AAI, and HM interpreted the results. All authors discussed the results and contributed to the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Assessment tests used in the *YogiTherapy* app to detect progress in mobility, disease activity, and function through Yoga exercises.
[DOCX File , 15 KB - formative_v6i6e34566_app1.docx]

References

1. Martins NA, Furtado GE, Campos MJ, Leitão JC, Filaire E, Ferreira JP. Exercise and ankylosing spondylitis with New York modified criteria: a systematic review of controlled trials with meta-analysis. Acta Reumatol Port 2014;39(4):298-308 [FREE Full text] [Medline: 25351868]
2. Sari İ, Öztürk MA, Akkoç N. Treatment of ankylosing spondylitis. Turk J Med Sci 2015;45(2):416-430 [FREE Full text] [doi: 10.3906/sag-1401-79] [Medline: 26084136]
3. Hilberdink B, van der Giesen F, Vliet Vlieland T, van Gaalen F, van Weely S. Supervised group exercise in axial spondyloarthritis: patients' satisfaction and perspective on evidence-based enhancements. Arthritis Care Res (Hoboken) 2020 Jun;72(6):829-837 [FREE Full text] [doi: 10.1002/acr.23892] [Medline: 30927518]
4. Jacquemin C, Servy H, Molto A, Sellam J, Foltz V, Gandjbakhch F, et al. Physical activity assessment using an activity tracker in patients with rheumatoid arthritis and axial spondyloarthritis: Prospective observational study. JMIR Mhealth Uhealth 2018 Jan 02;6(1):e1 [FREE Full text] [doi: 10.2196/mhealth.7948] [Medline: 29295810]
5. Azevedo R, Bernardes M, Fonseca J, Lima A. Smartphone application for rheumatoid arthritis self-management: cross-sectional study revealed the usefulness, willingness to use and patients' needs. Rheumatol Int 2015 Oct;35(10):1675-1685. [doi: 10.1007/s00296-015-3270-9] [Medline: 25903352]
6. Storm A, Bodmer T. Präventionsangebote. DAK Gesundheit. 2020. URL: https://www.dak.de/dak/leistungen/praeventionskurse-mit-ihrer-krankenkasse-2072756.html#/ [accessed 2021-10-01]
7. Storm A, Bodmer T. DAK Rücken@Fit - Online-Coaching für einen starken und gesunden Rücken. DAK Gesundheit. 2020. URL: https://www.dak.de/dak/leistungen/online-coaching-rueckenfit-2071724.html#/ [accessed 2021-10-01]
8. Apps gegen Rückenschmerzen. Krankenkassen Deutschland. 2021. URL: https://www.krankenkassen.de/gesetzliche-krankenkassen/leistungen-gesetzliche-krankenkassen/apps/ruecken/ [accessed 2021-10-01]
9. McDonald MT, Siebert S, Coulter EH, McDonald DA, Paul L. Level of adherence to prescribed exercise in spondyloarthritis and factors affecting this adherence: a systematic review. Rheumatol Int 2019 Feb;39(2):187-201 [FREE Full text] [doi: 10.1007/s00296-018-4225-8] [Medline: 30649569]
10. Knitza J, Simon D, Lambrecht A, Raab C, Tascilar K, Hagen M, et al. Mobile health usage, preferences, barriers, and eHealth literacy in rheumatology: patient survey study. JMIR Mhealth Uhealth 2020 Aug 12;8(8):e19661 [FREE Full text] [doi: 10.2196/19661] [Medline: 32678796]
11. Knitza J, Tascilar K, Messner EM, Meyer M, Vossen D, Pulla A, et al. German mobile apps in rheumatology: review and analysis using the Mobile Application Rating Scale (MARS). JMIR Mhealth Uhealth 2019 Aug 05;7(8):e14991 [FREE Full text] [doi: 10.2196/14991] [Medline: 31381501]
12. Kernder A, Morf H, Klemm P, Vossen D, Haase I, Mucke J, et al. Digital rheumatology in the era of COVID-19: results of a national patient and physician survey. RMD Open 2021 Feb;7(1):e001548 [FREE Full text] [doi: 10.1136/rmdopen-2020-001548] [Medline: 33622673]
13. Heidel A, Hagist C. Potential benefits and risks resulting from the introduction of health apps and wearables into the German statutory health care system: scoping review. JMIR Mhealth Uhealth 2020 Sep 23;8(9):e16444 [FREE Full text] [doi: 10.2196/16444] [Medline: 32965231]
14. Hoekstra F, Meeuwissen S. ASAS App. Assessment of SpondyloArthritis. URL: https://www.asas-group.org/education/asas-app/ [accessed 2021-03-01]
15. Zochling J. Measures of symptoms and disease status in ankylosing spondylitis: Ankylosing Spondylitis Disease Activity Score (ASDAS), Ankylosing Spondylitis Quality of Life Scale (ASQoL), Bath Ankylosing Spondylitis Disease Activity Index (BASDAI), Bath Ankylosing Spondylitis Functional Index (BASFI), Bath Ankylosing Spondylitis Global Score (BAS-G), Bath Ankylosing Spondylitis Metrology Index (BASMI), Dougados Functional Index (DFI), and Health Assessment Questionnaire for the Spondylarthropathies (HAQ-S). Arthritis Care Res (Hoboken) 2011 Nov;63 Suppl 11:S47-S58 [FREE Full text] [doi: 10.1002/acr.20575] [Medline: 22588768]
16. Exercise: Exercise matters - take charge - take action. National Axial Spondyloarthritis Society. URL: https://nass.co.uk/managing-my-as/exercise/ [accessed 2021-03-01]
17. O'Sullivan D, Wyke A. iAnkylosing Spondylitis. My health Apps. URL: http://myhealthapps.net/app/details/53/iAnkylosing-Spondylitis [accessed 2021-03-01]
18. O'Sullivan D, Wyke A. RheumaTrack. My Health Apps. URL: http://myhealthapps.net/app/details/205/RheumaTrack [accessed 2021-03-01]
19. Hoekstra F, Meeuwissen S. Asas App. Assessment of SpondyloArthritis. URL: https://www.asas-group.org/education/asas-app/ [accessed 2021-10-01]
20. Fitness, so vielfältig wie du. Gymondo. URL: https://www.gymondo.com/de/ [accessed 2021-10-01]
21. kaia health. URL: https://www.kaiahealth.com/de/ [accessed 2021-10-01]
22. Ward L, Stebbings S, Athens J, Cherkin D, David Baxter G. Yoga for the management of pain and sleep in rheumatoid arthritis: a pilot randomized controlled trial. Musculoskeletal Care 2018 Mar;16(1):39-47. [doi: 10.1002/msc.1201] [Medline: 28621011]
23. Gautam S, Kumar M, Kumar U, Dada R. Effect of an 8-week yoga-based lifestyle intervention on psycho-neuro-immune axis, disease activity, and perceived quality of life in rheumatoid arthritis patients: a randomized controlled trial. Front Psychol 2020 Sep 2;11:2259 [FREE Full text] [doi: 10.3389/fpsyg.2020.02259] [Medline: 32982898]

24. DRKS - German Clinical Trials Register. URL: https://www.drks.de/ [accessed 2022-05-17]
25. Schrepp M, Hinderks A, Thomaschewski J. Applying the user experience questionnaire (UEQ) in different evaluation scenarios. In: Proceedings of the 3rd International Conference on Design, User Experience, and Usability. Theories, Methods, and Tools for Designing the User Experience. 2014 Presented at: DUXU '14; June 22-27, 2014; Crete, Greece p. 383-392. [doi: 10.1007/978-3-319-07668-3_37]
26. Usage-share-of-operating-systems. Wikipedia - The Free Encyclopedia. URL: https://en.wikipedia.org/wiki/Usage-share-of-operating-systems [accessed 2021-03-01]
27. Lin Z, Gu J, He P, Gao J, Zuo X, Ye Z, et al. Multicenter validation of the value of BASFI and BASDAI in Chinese ankylosing spondylitis and undifferentiated spondyloarthropathy patients. Rheumatol Int 2011 Feb;31(2):233-238 [FREE Full text] [doi: 10.1007/s00296-009-1313-9] [Medline: 20012866]
28. van Weely SF, van Denderen CJ, van der Horst-Bruinsma IE, Nurmohamed MT, Dijkmans BA, Dekker J, et al. Reproducibility of performance measures of physical function based on the BASFI, in ankylosing spondylitis. Rheumatology (Oxford) 2009 Oct;48(10):1254-1260. [doi: 10.1093/rheumatology/kep190] [Medline: 19638455]
29. Irons K, Harrison H, Thomas A, Martindale J. Ankylosing Spondylitis (Axial Spondyloarthritis) - The Bath Indices. National Ankylosing Spondylitis Society (NASS). 2004. URL: https://nass.co.uk/wp-content/uploads/2018/09/Bath-Indices.pdf [accessed 2021-11-30]
30. Branaghan RJ, O'Brian JS, Hildebrand EA, Foster LB. Usability evaluation. In: Branaghan RJ, O'Brian JS, Hildebrand EA, Foster LB, editors. Humanizing Healthcare – Human Factors for Medical Device Design. Cham, Switzerland: Springer; 2021:69-96.
31. Schrepp M, Thomaschewski J, Hinderks A. Construction of a benchmark for the user experience questionnaire (UEQ). Int J Interact Multimed Artif Intell 2017;4(4):40-44. [doi: 10.9781/ijimai.2017.445]
32. Schrepp M, Pérez Cota MP, Gonçalves R, Hinderks A, Thomaschewski J. Adaption of user experience questionnaires for different user groups. Univ Access Inf Soc 2017 Aug;16(3):629-640. [doi: 10.1007/s10209-016-0485-9]
33. Schrepp M, Hinderks A, Thomaschewski J. UEQ excel data analysis tool. 2021. URL: https://www.ueq-online.org/ [accessed 2021-03-01]
34. Clement I, Lorenz A, Ulm B, Plidschun A, Huber S. Implementing systematically collected user feedback to increase user retention in a mobile app for self-management of low back pain: retrospective cohort study. JMIR Mhealth Uhealth 2018 Jun 06;6(6):e10422 [FREE Full text] [doi: 10.2196/10422] [Medline: 29875088]
35. Huber S, Priebe JA, Baumann KM, Plidschun A, Schiessl C, Tölle TR. Treatment of low back pain with a digital multidisciplinary pain treatment app: short-term results. JMIR Rehabil Assist Technol 2017 Dec 04;4(2):e11 [FREE Full text] [doi: 10.2196/rehab.9032] [Medline: 29203460]

Abbreviations

AS: axial spondyloarthritis
ASAS: Assessment of SpondyloArthritis
BASDAI: Bath Ankylosing Spondylitis Disease Activity Index
BASFI: Bath Ankylosing Spondylitis Functional Index
BASMI: Bath Ankylosing Spondylitis Metrology Index
DHA: digital health app
GUI: graphical user interface
TCR: task completion rate
UEQ: User Experience Questionnaire

Edited by T Leung; submitted 29.10.21; peer-reviewed by A Hueber, B Chaudhry, F Cotte, J Mucke, K Lim; comments to author 17.01.22; revised version received 02.02.22; accepted 30.03.22; published 03.06.22.

Please cite as:
Truong MT, Nwosu OB, Gaytan Torres ME, Segura Vargas MP, Seifer AK, Nitschke M, Ibrahim AA, Knitza J, Krusche M, Eskofier BM, Schett G, Morf H
A Yoga Exercise App Designed for Patients With Axial Spondylarthritis: Development and User Experience Study
JMIR Form Res 2022;6(6):e34566
URL: https://formative.jmir.org/2022/6/e34566
doi:10.2196/34566
PMID:35657655

Original Paper

Mobile Health App for Tuberculosis Screening and Compliance to Undergo Chest X-ray Examination Among Presumptive Cases Detected by the App in Myanmar: Usability Study

Kyaw Ko Ko Htet[1*], MSc, MD; Aye Nyein Phyu[2*], MD, MMedSc; Thandar Thwin[2*], MD; Virasakdi Chongsuvivatwong[3*], MD

[1]Department of Medical Research, Ministry of Health and Sports, Pyin Oo Lwin, Myanmar

[2]National Tuberculosis Programme, Department of Public Health, Ministry of Health and Sports, Mandalay, Myanmar

[3]Department of Epidemiology, Faculty of Medicine, Prince of Songkla University, Hat Yai, Thailand

[*]all authors contributed equally

Corresponding Author:
Virasakdi Chongsuvivatwong, MD
Department of Epidemiology
Faculty of Medicine
Prince of Songkla University
15 Kanjanavanich Road
Hat Yai, 90110
Thailand
Phone: 66 74451165
Fax: 66 74 429754
Email: cvirasak@medicine.psu.ac.th

Abstract

Background: In Myanmar, the use of a mobile app for tuberculosis (TB) screening and its operational effect on seeking TB health care have not been evaluated yet.

Objective: This study aims to report the usability of a simple mobile app to screen TB and comply with chest X-ray (CXR) examination of presumptive cases detected by the app.

Methods: A new "TB-screen" app was developed from a Google Sheet based on a previously published algorithm. The app calculates a TB risk propensity score from an individual's sociodemographic characteristics and TB clinical history and suggests whether the individual should undergo a CXR. The screening program was launched in urban slum areas soon after the COVID-19 outbreak subsided. A standard questionnaire was used to assess the app's usability rated by presumptive cases. Compliance to undergo CXR was confirmed by scanning the referral quick response (QR) code via the app.

Results: Raters were 453 presumptive cases detected by the app. The mean usability rating score was 4.1 out of 5. Compliance to undergo CXR examination was 71.1% (n=322). Active TB case detection among CXR compliances was 7.5% (n=24). One standard deviation (SD) increase in the app usability score was significantly associated with a 59% increase in the odds to comply with CXR (β=.464) after adjusting for other variables (P<.001).

Conclusions: This simple mobile app got a high usability score rated by 453 users. The mobile app usability score successfully predicted compliance to undergo CXR examination. Eventually, 24 (7.5%) of 322 users who were suspected of having TB by the mobile app were detected as active TB cases by CXR. The system should be upscaled for a large trial.

(JMIR Form Res 2022;6(6):e37779) doi:10.2196/37779

KEYWORDS

usability; mobile app; TB screening; chest X-ray compliance; mobile health; health application; risk score; tuberculosis; COVID-19

Introduction

In global practice, tuberculosis (TB) screening based on signs and symptoms is the first step in a case-finding strategy. Those having positive results should proceed to have their diagnosis confirmed with a chest X-ray (CXR) and sputum examination [1]. Previous analysis of TB screening data revealed that a TB risk propensity score calculated from sociodemographic and TB clinical variables could produce a better prediction of active TB than that based on TB signs and symptoms alone (80.6% vs 59.8% in sensitivity, 63.5% vs 67.2% in specificity, and 80.5% vs 63.7% in the area under the curve of a receiver operating characteristic curve) [2].

Based on these findings, a simple mobile health (mHealth) app was created and applied in an area close to where the score was developed. The app uses the statistics from Htet et al [2] to identify high-risk people who have a ≥0.5% probability to develop TB. They were then suggested to undergo CXR examination free of charge.

Standard mHealth app usability questionnaires have been developed and well tested [3]. For continuous improvement of the screening program, usability of the app needs to be assessed. In addition, to assess the operational effect of the app, it is important to follow up on the compliance of presumptive TB cases detected by the app and whether they would proceed to TB health center services for CXR examination as early as possible. In Myanmar, the use of a mobile app for TB screening and its effect on TB health services have not been assessed yet.

The aims of this study were (1) to assess the usability of our mHealth app perceived by presumptive TB cases, (2) to assess their compliance to proceed to undergo CXR examination, and (3) to determine the association between the usability of the app and compliance to undergo CXR examination.

Methods

Study Setting

This study was carried out in a low-income urban area of Mandalay City, which is densely populated and located in the central part of Myanmar [4-6]. Mandalay City has 7 township health departments providing TB services. Free medical services for TB diagnosis, including CXR and treatment, are available at TB health centers of the township health departments [7]. Routine TB care services are provided with precautions of COVID-19 prevention and control measures. For this study, 1 urban slum area was randomly selected from the list from each of 3 randomly selected townships by township medical officers and the regional TB coordinator.

Developing a Mobile Health App

The “TB-screen” app was developed by AppSheet's no-code app (Seattle, WA, USA) [8]. It was a collaborative effort between the Department of Medical Research, National TB Control Programme, Myanmar, and the Department of Epidemiology, Prince of Songkla University, Hat Yai, Thailand.

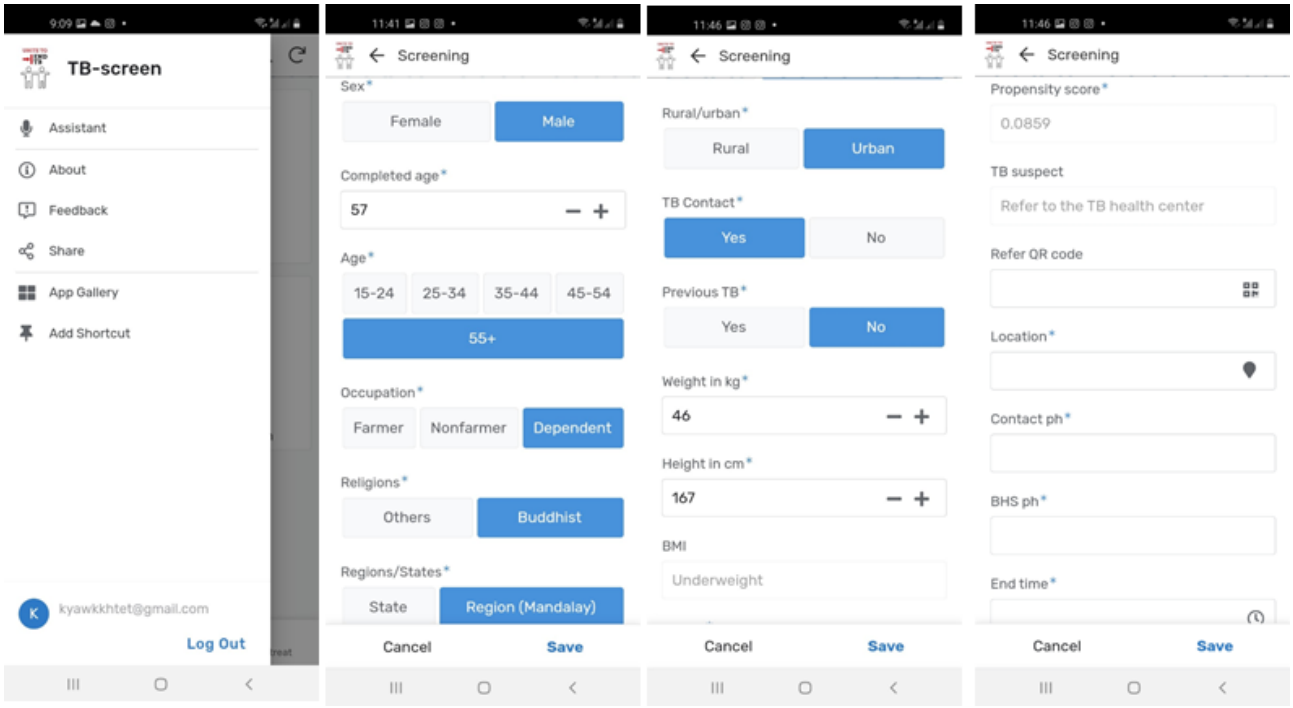
The app was built on the principal investigator’s computer directly from Google Sheet. It was designed to be used off-line on a mobile device running an Android operating system (Google, Inc., Mountain View, CA, USA). The app was installed in the mobile phones of health care providers who are responsible for TB screening. They were carefully trained on how to use the mHealth app. They invited a family household with their neighbors to join TB screening. After entering sociodemographic and TB clinical variables of an individual via the app, the TB risk propensity score is computed using the following formula [9]:

$$\text{TB risk propensity score} = \exp(BX)/[1 + \exp(BX)],$$

where B is a vector of regression coefficients and X is the matrix of sociodemographic and TB clinical covariates of the subject. Details of the variables and their related coefficients are shown in Multimedia Appendix 1. It was noted that the coefficient of the elderly age groups and male groups were higher than all other variables. The oldest age group (55+ years) had an $e^{2.4}$=11.0 times higher risk of developing TB than younger ones. The male group had an $e^{1.0}$=2.07 times higher risk of developing TB than the female group. These groups had high potential to be identified as presumptive TB cases via the mobile app. The TB risk propensity score is the value that measures the risk of developing TB [2].

With a selected propensity score cut-off level at ≥0.0052 (≥0.5% probability to develop TB), the test was taken as positive. The sensitivity is 80.6%, and the specificity is 63.5% [2]. A person with a positive test result, the presumptive TB case, is recommended to visit a TB health center for CXR examination. If they agree to go to the CXR center, a specific quick response (QR) code with a CXR referral form is scanned to the mHealth app (Figure 1). The presumptive TB case is given the CXR referral form to take to the CRX center. When the app is connected to the internet, the information collected is uploaded to the main server of data storage, with a notification to the corresponding TB health center. The data are updated again by the CXR center by scanning the referral QR code when the person visits and undergoes CXR examination. Figure 1 shows a screenshot of the TB-screen app.

Figure 1. Screenshot of the TB-screen app. TB: tuberculosis.



Ethical Considerations

The study was approved by the Institutional Ethics Committee of the Faculty of Medicine, Prince of Songkla University, Hat Yai, Thailand (REC:63-074-18-1), and the Institutional Review Board of the Department of Medical Research, Myanmar (IRB00008835).

Screening Procedures

The screening process was carried out during November 2020-January 2021. This was after the COVID-19 outbreak substantially subsided and lockdowns and travel restrictions were eased. A trained local health care provider who is responsible for providing both TB and COVID-19 care in the assigned community performed TB screening via the app, following COVID-19 prevention and control measures. For each community, a family household with neighbors was approached for TB screening. Only 4 or 5 family households were approached, and at most 5-6 presumptive TB cases were referred for CXR examination per day to avoid overloading at the CXR center. The health care providers had been vaccinated with the complete dose of the COVID-19 vaccine. The presumptive TB cases were provided with face masks and face shields to be used during the interview and when they went to the CXR center. The presumptive TB cases with suspected COVID-19 were referred to the community fever clinic center at the respective township health department for COVID-19 investigation. Those who tested negative for COVID-19 proceeded to the CXR center for CXR examination.

At the CXR center, there was a separate waiting room for the presumptive TB cases from the same community, with the chairs arranged with social distancing measures. One CXR technician was assigned for the CXR examination during the project. After CXR examination was performed, the presumptive TB cases returned home without waiting for CXR results. They were informed about the CXR results by the health care provider who performed TB screening. The presumptive TB cases needed a single visit for CXR examination, and their remaining diagnostic and treatment procedures were aided by the health care provider and local health volunteers. The presumptive TB cases with abnormal suggestion of TB on CXR were confirmed as active TB cases by performing Gene Xpert *Mycobacterium tuberculosis* complex and resistance to rifampin (MTB/RIF) examination.

Data Collection

During screening, informed consent for those >18 years old or guardian's assent for those 15-18 years old was obtained for each participant. The screened participants were informed about survey processes using the app, and they were interviewed about their background, sociodemographic, and TB clinical characteristics; perceived susceptibility to developing TB; perceived benefits of TB screening; and perceived harms of TB screening. After collecting the variables related to the TB risk propensity score via the app, information about usability of the mHealth app was collected. Point coordinate data for location of participants were taken by the app for further use to calculate the distance between their residences and the TB health center for CXR. Accessibility methods to the CXR center were also recorded.

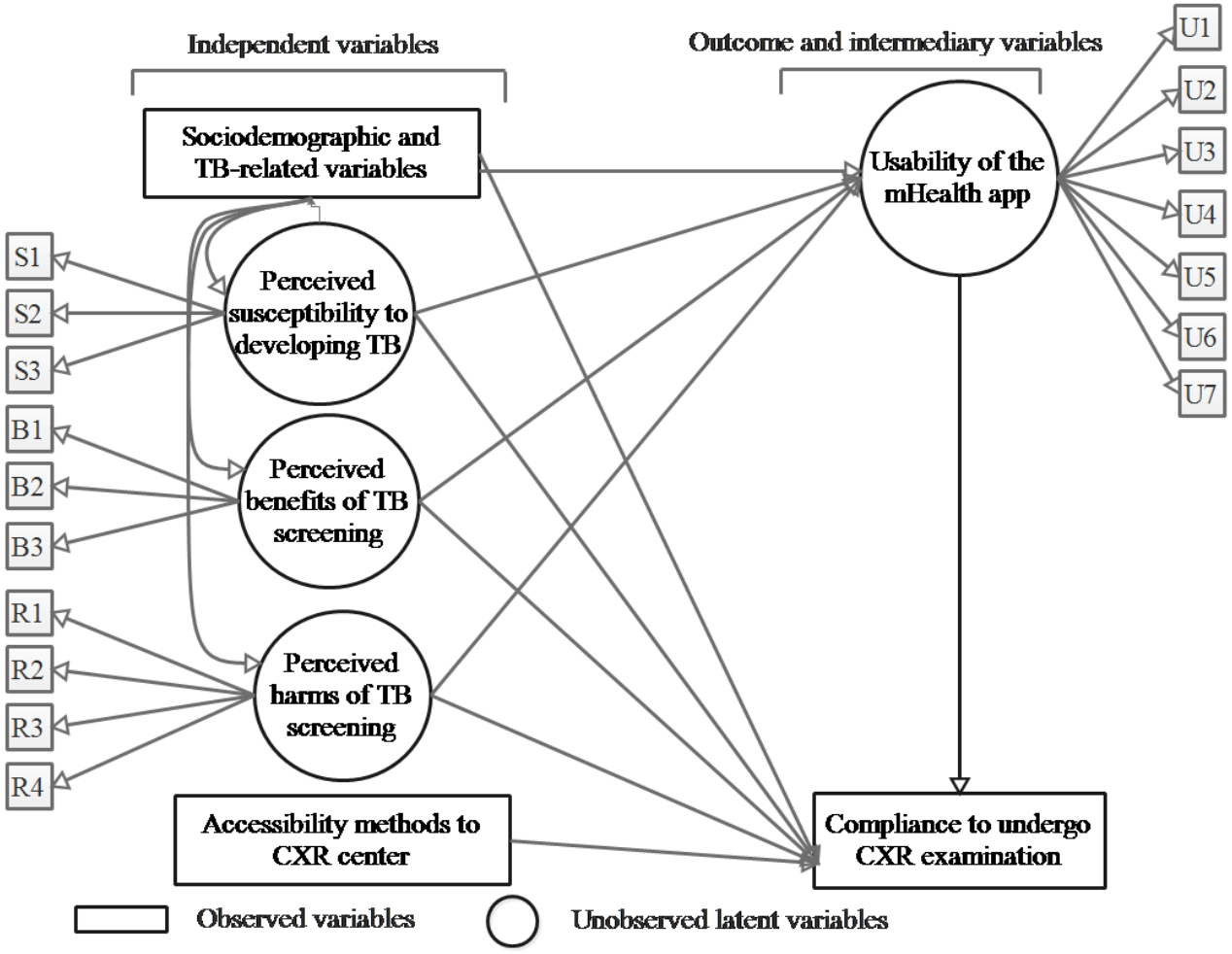
Statistical Analysis

Proposed Model of the Analysis

Figure 2 illustrates the proposed major relationships among the variables and their components. The usability of the app was an intermediary latent variable influenced by background, sociodemographic, and clinical variables, which also influenced 3 latent variables, namely perceived susceptibility to developing TB, perceived benefits of TB screening, and perceived harms of TB screening. All these variables, including usability of the

app, influenced compliance to CXR examination. Finally, compliance to CXR examination was also influenced by accessibility methods to a CXR center.

Figure 2. Baseline model to determine the association between usability of the app and compliance to undergo CXR examination and their influencing factors. Small boxes with U, S, B, and R denote items measured for the respective latent variable. CXR: chest X-ray; TB: tuberculosis.



Outcome Variable

The primary outcome was compliance of presumptive TB cases detected by the app to undergo CXR examination within 1-7 days of TB screening.

Intermediary Variable

The main intermediary variable was the usability of the mHealth app, which was a latent variable. It was measured by using the standard mHealth app usability questionnaire (U1-U7 items, Cronbach α=.9) [3].

Independent Variables

The sociodemographic and TB-related variables included marital status, education, family income per month (US $), the TB risk propensity score, TB signs and symptoms, and knowledge of TB.

The TB risk propensity score was derived from age, gender, occupation, religion, area of residence, administrative division, contact with a known TB case, previous history of TB, and BMI (kg/m^2) [2].

TB signs and symptoms included cough, hemoptysis, recent loss of weight, chest pain, and fever within the previous 1 month. There were 5 TB knowledge questions adopted from the World Health Organization (WHO) TB survey, which asked about TB signs and symptoms, persons who are at high risk of developing TB, transmission, diagnosis methods, and cure [10]. A correct answer was assigned a score of 1, and 0 otherwise.

Items related to perceived susceptibility to developing TB, perceived benefits of TB screening, and perceived harms of TB screening via the app were constructed using Champion and Skinner's variable definitions [11]. The perceived susceptibility to developing TB was assessed by 3 items (S1-S3); for example, "You are at high risk of TB infection." The perceived benefits of TB screening were constructed by 3 items (B1-B3); for example, "TB screening is good for your health." The perceived harms of TB screening were identified by 4 items (R1-R4); for example, "You are afraid of developing TB."

The variables of accessibility methods to accessing a CXR center were availability to go to the center during clinic opening hours, having one's own vehicle to go to the CXR center, and traveling distance (km) to the CXR center.

Confirmatory Factor Analysis

Confirmatory factor analysis (CFA) was performed to verify the fit of the observed items, internal consistency, and discriminant validity to each latent variable [12-15].

Structural Analysis of the Causal Pathway in Structural Equation Modeling

The proposed model was analyzed by structural equation modeling (SEM). Based on the construction illustrated in Figure 2, the structural analysis of SEM was used to examine whether our baseline proposed model was acceptable and a good fit [16]. The independent variables having a *P* value ≥.05 were dropped each time until the goodness-of-fit was acceptable. The model is acceptable when the Tucker-Lewis fit index (TLI)≥0.95, the ratio of the chi-square statistic to its degrees of freedom is close to 1, cumulative fit index (CFI)≥0.95, root-mean-square error of approximation (RMSEA)≤0.05, and standardized root-mean-square residual (SRMR)≤0.08 [12-14].

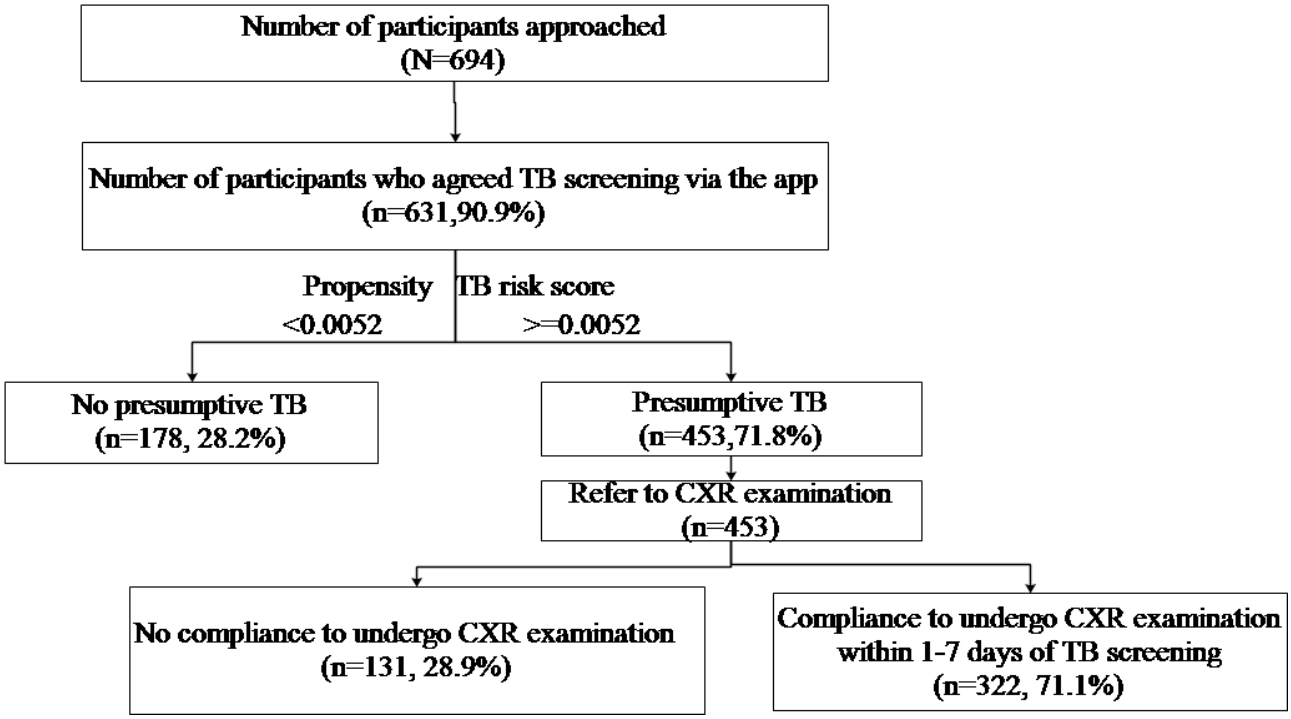
In the final fitted SEM model, the standardized estimate (β), regression coefficient, and standard error (SE) of observed variables and latent variables associated with usability of the app and compliance to undergo CXR examination were calculated. The data were analyzed using R version 4.0.0 (R Foundation for Statistical Computing) [17].

Results

Subject Classification

Figure 3 summarizes the overall results of subject classification based on the TB risk propensity score among the screened participants and compliance to undergo CXR examination among the presumptive TB cases. Of 694 participants approached, 631 (90.9%) agreed to be screened via the app. Among those screened, 453 (71.8%) were identified as presumptive TB cases by the app and were suggested to visit a TB health center for a CXR appointment.

Figure 3. Flowchart of participants under TB screening until visit for CXR examination. CXR: chest X-ray; TB: tuberculosis.



Compliance to Undergo CXR Examination

As shown in Figure 3, of 453 presumptive TB cases, 322 (71.1%) complied to undergo CXR examination within 1-7 days of TB screening.

Characteristics of Presumptive TB Cases Detected by the App (N=453)

Table 1 shows the background characteristics of the 453 presumptive TB cases. Males and the elderly had a high proportion, accounting for 251 (55.4%) and 150 (33.1%) participants, respectively. The mean (SD) age was 46.1 (15.0) years. The median TB risk propensity score was 0.01 (IQR 0.0058-0.022). A high risk score indicated an increase in the risk of developing TB. Over half (n=258, 57.0%) had no TB signs or symptoms. Knowledge of TB was high, with a mean (SD) score of 5.9 (1.4) out of 8. The cases had less perceived susceptibility to developing TB, with a mean (SD) score of 2.7 (1.0) out of 5, but more perceived harms of TB screening, with a mean (SD) score of 3.0 (1.1) out of 5.

Table 1. Characteristics of the presumptive TB[a] cases (N=453).

Variables and their description	Value
Marital status, n (%)	
Married	403 (89.0)
Single	50 (11.0)
Education, n (%)	
None	25 (5.5)
Primary school	220 (48.6)
Secondary school	101 (22.3)
Middle school	71(15.7)
High school and above	36 (7.9)
Family income per month (US $), n (%)	
≤80	112 (24.7)
81-240	320 (70.6)
241-400	17 (3.8)
>400	4 (0.9)
Gender, n (%)	
Female	202 (44.6)
Male	251 (55.4)
Age (years), n (%)	
15-24	29 (6.4)
25-34	82 (18.1)
35-44	106 (23.4)
45-54	86 (19.0)
+55	150 (33.1)
Age (years), mean (SD)	46.1 (15.0)
Religion, n (%)	
Buddhist	438 (96.7)
Others	15 (3.3)
Occupation, n (%)	
Dependent	218 (48.1)
Farmer	21 (4.7)
Nonfarmer	214 (47.2)
Contact with a known TB case, n (%)	
No	291 (64.2)
Yes	162 (35.8)
Previous history of TB, n (%)	
No	353 (77.9)
Yes	100 (22.1)
TB signs and symptoms, n (%)	
Yes	195 (43.0)
No	258 (57.0)
BMI (kg/m^2), mean (SD)	19.8 (3.3)

Variables and their description	Value
TB risk propensity score, median (IQR)	0.01 (0.0058-0.022)
Knowledge of TB (range 0-8), mean (SD)	5.9 (1.4)
Perceived susceptibility to developing TB[b], mean (SD)	2.7 (1.0)
Perceived benefits of TB screening[b], mean (SD)	4.4 (0.6)
Perceived harms of TB screening[b], mean (SD)	3.0 (1.1)
Available to go to CXR[c] center during clinic opening hours, n (%)	
Yes	289 (63.8)
No	164 (36.2)
Have an own vehicle to go to the CXR center, n (%)	
Yes	93 (20.5)
No	360 (79.5)
Distance to CXR center (km), n (%)	
≤10	211 (46.6)
>10	242 (53.4)

[a]TB: tuberculosis.

[b]Latent variables.

[c]CXR: chest X-ray.

Usability of the Mobile Health App

Table 2 shows the usability scores of the mHealth app rated by the presumptive TB cases. All items had an average rating score of >3. The overall rating on the usability of the mHealth app was favorable, with a mean (SD) score of 4.1 (1.1) out of 5, indicating a high level of usability rated by the users.

Table 2. Usability of the mHealth[a] app by the presumptive TB[b] detected by the app (N=453).

Item	Strongly disagree	Disagree	Neutral	Agree	Strongly agree	Mean (SD)	Cronbach α=.941
	1	2	3	4	5		
U1: The mobile app improves your access to TB health care services.	0	61	31	181	180	4.0 (1.0)	N/A[c]
U2: The mobile app makes it convenient for you to communicate with your health care provider.	0	79	14	145	215	4.1 (1.1)	N/A
U3: By using the mobile app in TB screening, you have many more opportunities to interact with the health care provider.	0	80	13	130	230	4.2 (1.1)	N/A
U4: You feel confident that any information you received from the mobile app.	0	51	44	145	213	4.0 (1.0)	N/A

[a]mHealth: mobile health.

[b]TB: tuberculosis.

[c]N/A: not applicable.

Confirmatory Factor Analysis

In Table 3, the final CFA showed a good fit after removing 3 items from the usability of the app (U5, U6, and U7) and 2 items from the perceived harms of TB screening (R3 and R4) with factor loadings of <0.3 in the baseline CFA. Each latent variable was rated using a 5-point Likert scale ranging from 1 for "totally disagree" to 5 for "totally agree."

Table 4 shows that the composite reliability was higher than the average variance extracted and the average variance extracted was >0.5. In addition, the square root of the average variance extracted was higher than the correlations of the latent variables under analysis. Thus, the latent variables had both good consistency and discriminant validity.

Table 3. CFA[a] of latent variables.

Items	Baseline			Final		
	Factor loading	Cronbach α	Model fit ($\chi^2_{(113)}$=1078.3, *P*<.001, CFI[b]=0.797, TLI[c]=0.756, RMSEA[d]=0.137, SRMR[e]=0.099)	Factor loading	Cronbach α	Model fit ($\chi^2_{(47)}$=74.1, *P*=.01, CFI=0.993, TLI=0.99, RMSEA=0.036, SRMR=0.026)
Usability of the mHealth[f] app	N/A[g]	.7919	N/A	N/A	.941	N/A
U1: The mobile app improves your access to TB[h] health care services.	0.872	N/A	N/A	0.872	N/A	N/A
U2: The mobile app makes it convenient for you to communicate with your health care provider.	0.764	N/A	N/A	0.764	N/A	N/A
U3: By using the mobile app in TB screening, you have many more opportunities to interact with the health care provider.	0.953	N/A	N/A	0.953	N/A	N/A
U4: You feel confident that any information you received from the mobile app.	0.961	N/A	N/A	0.961	N/A	N/A
U5: The app is useful for improving your health and well-being.	0.096	N/A	N/A	N/A	N/A	N/A
U6: You feel comfortable communicating with your health care provider using the app.	0.018	N/A	N/A	N/A	N/A	N/A
U7: The app helps you manage your health effectively.	0.058	N/A	N/A	N/A	N/A	N/A
Perceived susceptibility to developing TB	N/A	.921	N/A	N/A	.921	N/A
S1: You are at high risk of TB infection.	0.936	N/A	N/A	0.936	N/A	N/A
S2: You are probably infected with TB with or without having TB signs or symptoms.	0.850	N/A	N/A	0.850	N/A	N/A
S3: You are the most possible person to be infected with TB among all family members.	0.894	N/A	N/A	0.894	N/A	N/A
Perceived benefits of TB screening	N/A	.730	N/A	N/A	.730	N/A
B1: This screening tool is convenient to identify TB early.	0.712	N/A	N/A	0.712	N/A	N/A
B2: If results of the screening are positive, you can access a TB health center for early TB diagnosis.	0.877	N/A	N/A	0.877	N/A	N/A
B3: TB screening is good for your health.	0.502	N/A	N/A	0.502	N/A	N/A
Perceived harms of TB screening	N/A	.608	N/A	N/A	.926	N/A
R1: You are afraid of developing TB.	0.953	N/A	N/A	0.978	N/A	N/A
R2: You are afraid of suffering social stigma due to TB.	0.905	N/A	N/A	0.882	N/A	N/A
R3: Screening by using the mobile app protects your privacy.	0.159	N/A	N/A	N/A	N/A	N/A
R4: Screening by using the mobile app keeps your personal information confidential.	–0.033	N/A	N/A	N/A	N/A	N/A

[a]CFA: confirmatory factor analysis.

[b]CFI: comparative fit index.

[c]TLI: Tucker-Lewis index.

[d]RMSEA: root-mean-square error of approximation.

[e]SRMR: standardized root-mean-square residual.

[f]mHealth: mobile health.

[g]N/A: not applicable.

[h]TB: tuberculosis.

Table 4. Internal consistency and discriminant validity of latent variables in the fitted CFA[a].

Latent variables	Items	Composite reliability	Average variance extracted	Correlation coefficients			
				Usability of the mHealth[b] app	Perceived susceptibility to developing TB[c]	Perceived benefits of TB screening	Perceived harms of TB screening
Usability of the mHealth app	4	0.896	0.764	1	N/A[d]	N/A	N/A
Perceived susceptibility to developing TB	3	0.922	0.797	(–0.042)	1	N/A	N/A
Perceived benefits of TB screening	3	0.754	0.521	(0.058)	(0.001)	1	N/A
Perceived harms of TB screening	2	0.929	0.867	(–0.24)	(–0.088)	(0.01)	1

[a]CFA: confirmatory factor analysis.

[b]mHealth: mobile health.

[c]TB: tuberculosis.

[d]N/A: not applicable.

Structural Analysis of the Causal Pathway in SEM

As shown in Table 5, the final SEM model was acceptable (TLI>0.95) and showed a better fit than the baseline proposed model (CFI=0.955 vs 0.922).

Figure 4 shows the standardized coefficients (β) of variables in the structural analysis of the causal pathway in the final SEM.

In the final SEM, the usability of the mHealth app and compliance to undergo CXR examination were not associated with marital status, education, perceived susceptibility to developing TB, perceived benefits of TB screening, and variables of accessibility methods to the CXR center. For clarity, these variables are not shown in Figure 4.

The usability of the mHealth app was significantly associated with compliance to undergo CXR examination. As β=.464, 1 standard deviation increase in the usability score of the app would be associated with $e^{0.464}$=1.59-fold odds or 59% increase to comply to undergo CXR examination.

The other significant factors associated with compliance were a high TB risk propensity score and high TB knowledge. In contrast, those with a lack of TB signs and symptoms and with perceived harms of TB screening were less likely to comply to undergo CXR examination.

Similarly, having high TB knowledge favored the usability of the app, while a lack of TB signs and symptoms and having perceived harms of TB screening predisposed individuals to be less favorable toward the usability of the app.

Table 5. Comparison of baseline and final SEM[a] models.

SEM comparison	χ^2 (*df*)	*P* value	CFI[b]	TLI[c]	RMSEA[d] (90% CI)	SRMR[e]
Baseline	255.6 (143)	<.001	0.922	0.957	0.042 (0.033-0.05)	0.04
Final	52.9 (34)	.02	0.955	0.972	0.035 (0.014-0.053)	0.025

[a]SEM: structural equation modeling.

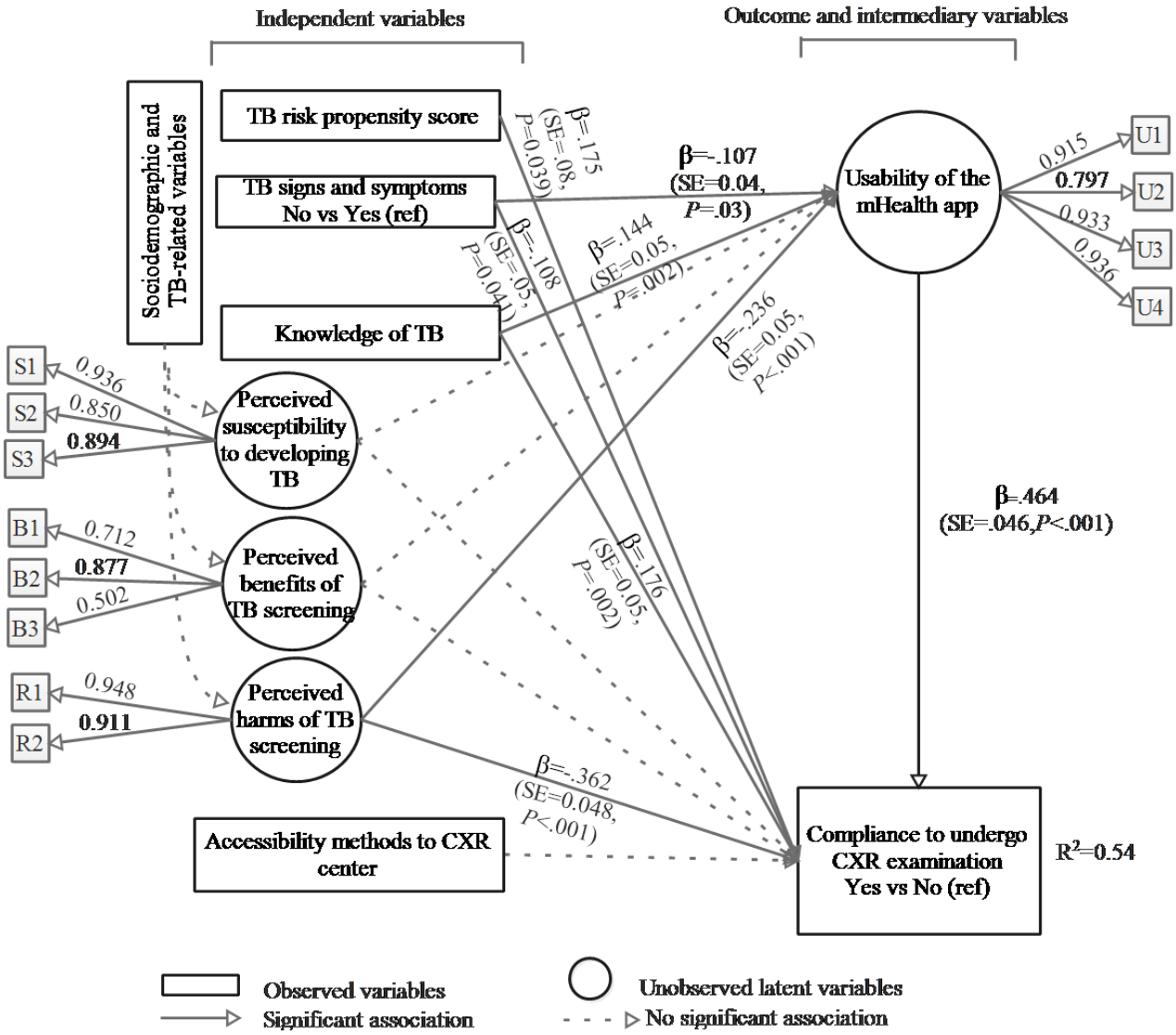
[b]CFI: comparative fit index.

[c]TLI: Tucker-Lewis index.

[d]RMSEA: root-mean-square error of approximation.

[e]SRMR: standardized root-mean-square residual.

Figure 4. Standardized coefficients (β) of variables in structural analysis of the causal pathway in the final fitted SEM. Note: Nonsignificant variables from sociodemographic variables, TB-related variables, and accessibility methods to CXR centers were dropped from the final SEM. Small boxes with U, S, B, and R denote items measured for the respective latent variable. CXR: chest X-ray; SEM: structural equation modeling; TB: tuberculosis.



Active TB Case Detection Among Presumptive TB Cases Who Complied to Undergo CXR Examination

Of 453 presumptive TB cases detected by the app, 322 (71.1%) complied to undergo CXR examination. Of them, 56 (17.4%) were identified as showing an abnormal suggestion of TB in the CXR and continued for Gene Xpert MTB/RIF examination. The total active TB case detection was 24 (7.5%) of 322 presumptive TB cases.

Discussion

Principal Findings

This is the first study in Myanmar using a mobile app for screening for TB and assessing its impact on seeking TB health care services. This mHealth app was rated by community members to have a good usability score. A significant proportion of the presumptive TB cases were identified via the app. Nearly three-quarters of the presumptive TB cases identified by the app proceeded to undergo CXR examination, and 7.5% showing CXR compliance were diagnosed as active TB cases. The usability of the mHealth app was the strongest predictor associated with the compliance to undergo CXR examination. Low predicting variables included TB signs and symptoms, perceived harms of TB screening, the TB risk propensity score, and knowledge of TB.

In this study, a high number of presumptive TB cases were identified via the mHealth app and TB health centers were notified. This could be explained by the high percentages (55.4% and 33.1%) of subjects under screening being male and elderly, respectively. The setting of the app gave a high score to these groups. In addition, nearly three-quarters of the presumptive TB cases identified by the app complied to undergo CXR examination. It was relatively high compared to a previous study in Myanmar [18]. Eventually, 24 (7.5%) of 322 users who were suspected by the app as having TB were detected as active TB cases by CXR. This proportion was missing active TB cases in the community that were early detected by the mobile app. However, this achievement would not be possible without the compliance to undergo CXR examination by the presumptive

TB cases detected by the app. Completion of this step would lead to early detection of TB cases, achieving the WHO-recommended "universal access to test and treat process" [19].

During implementation of the app, the participants rated the usability of the mHealth app highly and a significant proportion of presumptive TB cases were detected via the app. Many studies have revealed patients' acceptance of using mHealth apps in TB care to improve health outcomes [20,21]. In addition, the favorable rating for the usability of the mHealth app had a positive association with compliance to undergo CXR examination. In this study, the mobile app for TB screening was integrated into a pathway to TB diagnosis by notifying the CXR center of screened positive cases. This integration might be the reason for the high association of these two variables. A similar improvement in TB referrals by engaging the presumptive TB cases in the health sector via the app was shown in a study in India [22].

In this study, presumptive TB cases without TB signs or symptoms and those who had perceived harms of TB screening gave a less favorable rating to the usability of the app and had poor compliance to undergo CXR examination. A review of national TB prevalence surveys in Asia highlighted that those who screened negative for TB symptoms were less likely to seek TB care until symptoms worsened [23]. Many studies have revealed that having perceived harms of TB screening has a negative impact on TB screening and subsequent diagnostic procedures [24-26]. Having knowledge of TB can increase acceptance of patients with TB and encourage subsequent screening and diagnostic processes [27].

Strengths and Limitations

Compared to similar studies, this mHealth app calculated the TB risk propensity score instead of using TB signs and symptoms [22,28-30]. Although the TB risk propensity score has better TB prediction than TB signs and symptoms, there is a limitation in applying it in the real field, probably because of its complexity to calculate [31,32]. However, our mobile app overcomes this gap.

In addition, a review of a national TB prevalence survey in Asia (1990-2012) revealed that 40%-60% of active TB cases are missed by screening for routine TB signs and symptoms, because this proportion of patients are asymptomatic and not identified as presumptive TB cases [23]. In Myanmar, the TB case detection rate by screening for routine TB signs and symptoms was 69%-77% in 2017-2019 [33-35]. This highlighted that nearly one-fourth of active TB cases were missing, either not notified to the national TB program or not diagnosed and treated. Therefore, using a mobile app that calculates the TB risk score is a potential new approach to identifying missing presumptive TB and active TB cases that would not be detected by TB signs and symptoms only.

As strengths, the mobile app was used to calculate the TB risk propensity score so misclassification bias to identify the TB suspects was reduced. Real-time data entry was performed for the CXR examination date so that data error was less likely in assessing compliance. The app is simple to develop, provided that existing TB survey databases can be used to calculate the TB risk propensity score for the population under study. This idea can be adapted to many low-resource countries.

However, a high level of compliance to undergo CXR examination in this study was probably due to worry about long COVID among the population after a serious COVID-19 outbreak recently subsided [36]. In addition to TB detection, they might expect CXR to detect the residual effects of COVID-19 and thus they had high compliance. As a limitation, high compliance to undergo CXR examination may suggest that the TB screening program could be done better immediately after a COVID-19 outbreak or its generalizability should be confirmed when the COVID-19 situation is no more perceived as a serious threat.

Conclusion

The simple mobile app we developed got a high usability score by 453 users. The mobile app usability score successfully predicted compliance to undergo CXR examination. Eventually, 24 of 322 users who were suspected by the mobile app as having TB were detected as active TB cases by CXR. The system should be upscaled for a large trial.

Acknowledgments

This study is part of the thesis of KKKH in partial fulfillment of the requirement of a degree of doctor of philosophy in epidemiology at Prince of Songkla University, Hat Yai, Songkhla, Thailand. The research was supported by the Fogarty International Center and the National Institute of Allergy and Infectious Disease of the National Institutes of Health (NIH) for the project titled "TB/MDR-TB Research Capacity Building in Low- and Middle-Income Countries in Southeast Asia" (Award D43TW009522) awarded to VC. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. This research work was also cosupported by the sponsorship of Discipline of Excellence (Epidemiology), Prince of Songkla University.

The authors wish to thank the study participants for their contribution to the research as well as township medical officers at Mandalay City and research assistants from the Department of Medical Research (Pyin Oo Lwin branch). Our sincere thanks are directed to the National Tuberculosis Programme, Department of Public Health, Ministry of Health and Sports, Nay Pyi Taw, and the Department of Medical Research, Ministry of Health and Sports, for allowing us to conduct this research.

Conflicts of Interest

The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Multimedia Appendix 1
Regression coefficient (β) of significant sociodemographic and TB clinical history variables to predict bacteriologically confirmed TB. TB: tuberculosis.
[DOCX File , 15 KB - formative_v6i6e37779_app1.docx]

References

1. World Health Organization. Systematic Screening for Active Tuberculosis: An Operational Guide. URL: https://apps.who.int/iris/bitstream/handle/10665/181164/9789241549172_eng.pdf [accessed 2022-06-02]
2. Htet KKK, Chongsuvivatwong V, Aung ST. Sensitivity and specificity of tuberculosis signs and symptoms screening and adjunct role of social pathology characteristics in predicting bacteriologically confirmed tuberculosis in Myanmar. Trop Med Health 2021 Jan 07;49(1):3 [FREE Full text] [doi: 10.1186/s41182-020-00292-x] [Medline: 33407932]
3. Zhou L, Bao J, Setiawan IMA, Saptono A, Parmanto B. The mHealth App Usability Questionnaire (MAUQ): development and validation study. JMIR Mhealth Uhealth 2019 Apr 11;7(4):e11500 [FREE Full text] [doi: 10.2196/11500] [Medline: 30973342]
4. Ministry of Health and Sports (Myanmar). National Tuberculosis Program Myanmar: Annual Report 2015. URL: https://www.aidsdatahub.org/sites/default/files/resource/national-tb-programme-myanmar-annual-report-2015.pdf [accessed 2022-06-02]
5. Thin Zaw PP, Liabsuetrakul T, McNeil E, Htay TT. Gender differences in exposure to SRH information and risky sexual debut among poor Myanmar youths. BMC Public Health 2013 Dec 05;13(1):1122 [FREE Full text] [doi: 10.1186/1471-2458-13-1122] [Medline: 24304552]
6. Department of Population. Census Atlas Myanmar: The 2014 Myanmar Population and Housing Census. URL: https://myanmar.unfpa.org/en/publications/census-atlas-myanmar [accessed 2019-01-18]
7. Ministry of Health. Health in Myanmar, 2014. URL: http://themimu.info/sites/themimu.info/files/documents/Report_Health_in_Myanmar_2014_MOH.pdf [accessed 2022-06-02]
8. Google. AppSheet: The Fastest Way to Build Apps and Automate Work. URL: https://www.appsheet.com/ [accessed 2022-06-02]
9. Columbia University Mailman School of Public Health. Propensity Score Analysis. URL: https://www.mailman.columbia.edu/research/population-health-methods/propensity-score [accessed 2022-06-02]
10. World Health Organization. Advocacy, Communication and Social Mobilization for TB Control: A Guide to Developing Knowledge, Attitude and Practice Surveys. URL: https://apps.who.int/iris/handle/10665/43790 [accessed 2022-06-02]
11. Skinner CS, Tiro J, Champion VL. The health belief model. In: Glanz K, Rimer BK, Viswanath KV, editors. Health Behavior: Theory, Research, and Practice. San Francisco: Jossey-Bass/Wiley; 2015:75-94.
12. Brown TA. Confirmatory Factor Analysis for Applied Research. Second Edition. New York, NY: Guilford Press; 2015.
13. Portela DMP. Contributo das técnicas de análise fatorial para o estudo do programa “ocupação científica de jovens nas férias”. Portugal: Universidade Aberta; 2012.
14. Pavlov G, Maydeu-Olivares A, Shi D. Using the standardized root mean squared residual (SRMR) to assess exact fit in structural equation models. Educ Psychol Meas 2021 Feb 08;81(1):110-130 [FREE Full text] [doi: 10.1177/0013164420926231] [Medline: 33456064]
15. Fornell C, Larcker DF. Structural equation models with unobservable variables and measurement error: algebra and statistics. J Mark Res 2018 Nov 28;18(3):382-388. [doi: 10.1177/002224378101800313]
16. Kupek E. Beyond logistic regression: structural equations modelling for binary variables and its application to investigating unobserved confounders. BMC Med Res Methodol 2006 Mar 15;6(1):13 [FREE Full text] [doi: 10.1186/1471-2288-6-13] [Medline: 16539711]
17. R Core Team. R: A Language and Environment for Statistical Computing. 2017. URL: https://www.R-project.org/ [accessed 2022-06-02]
18. Thin PP, Mongkolchati A, Laosee O. Delays in care seeking treatment among persons with new pulmonary tuberculosis in Mandalay District, Central Myanmar. J Public Health Dev 2020;18(1):9-19.
19. World Health Organization. Early Detection of Tuberculosis: An Overview of Approaches, Guidelines and Tools. URL: https://apps.who.int/iris/handle/10665/70824 [accessed 2022-06-02]
20. Hirsch-Moverman Y, Daftary A, Yuengling KA, Saito S, Ntoane M, Frederix K, et al. Using mHealth for HIV/TB treatment support in Lesotho. JAIDS 2017;74:S37-S43. [doi: 10.1097/qai.0000000000001202]
21. Kumar AA, De Costa A, Das A, Srinivasa G, D'Souza G, Rodrigues R. Mobile health for tuberculosis management in South India: is video-based directly observed treatment an acceptable alternative? JMIR Mhealth Uhealth 2019 Apr 03;7(4):e11687 [FREE Full text] [doi: 10.2196/11687] [Medline: 30942696]

22. Chadha S, Trivedi A, Nagaraja SB, Sagili K. Using mHealth to enhance TB referrals in a tribal district of India. Public Health Action 2017 Jun 21;7(2):123-126 [FREE Full text] [doi: 10.5588/pha.16.0080] [Medline: 28695085]
23. Onozaki I, Law I, Sismanidis C, Zignol M, Glaziou P, Floyd K. National tuberculosis prevalence surveys in Asia, 1990-2012: an overview of results and lessons learned. Trop Med Int Health 2015 Sep 07;20(9):1128-1145 [FREE Full text] [doi: 10.1111/tmi.12534] [Medline: 25943163]
24. Courtwright A, Turner AN. Tuberculosis and stigmatization: pathways and interventions. Public Health Rep 2010 Jul 01;125(4_suppl):34-42. [doi: 10.1177/00333549101250s407]
25. Jaramillo E. Pulmonary tuberculosis and health-seeking behaviour: how to get a delayed diagnosis in Cali, Colombia. Trop Med Int Health 1998 Feb;3(2):138-144 [FREE Full text] [doi: 10.1046/j.1365-3156.1998.00202.x] [Medline: 9537276]
26. Rubel AJ, Garro LC. Social and cultural factors in the successful control of tuberculosis. Public Health Rep 1992;107(6):626-636 [FREE Full text] [Medline: 1454974]
27. Idris NA, Zakaria R, Muhamad R, Nik Husain NR, Ishak A, Wan Mohammad WMZ. The effectiveness of tuberculosis education programme in Kelantan, Malaysia on knowledge, attitude, practice and stigma towards tuberculosis among adolescents. Malays J Med Sci 2020 Dec 29;27(6):102-114 [FREE Full text] [doi: 10.21315/mjms2020.27.6.10] [Medline: 33447138]
28. Ha YP, Littman-Quinn R, Antwi C, Seropola G, Green RS, Tesfalul MA, et al. A mobile health approach to tuberculosis contact tracing in resource-limited settings. Stud Health Technol Inform 2013;192:1188-1121. [doi: 10.3233/978-1-61499-289-9-1188] [Medline: 27668973]
29. Velayutham B, Thomas B, Nair D, Thiruvengadam K, Prashant S, Kittusami S, et al. The usefulness and feasibility of Mobile Interface in Tuberculosis Notification (MITUN) voice based system for notification of tuberculosis by private medical practitioners--a pilot project. PLoS One 2015 Sep 16;10(9):e0138274 [FREE Full text] [doi: 10.1371/journal.pone.0138274] [Medline: 26376197]
30. Maraba N, Hoffmann CJ, Chihota VN, Chang LW, Ismail N, Candy S, et al. Using mHealth to improve tuberculosis case identification and treatment initiation in South Africa: results from a pilot study. PLoS One 2018 Jul 3;13(7):e0199687 [FREE Full text] [doi: 10.1371/journal.pone.0199687] [Medline: 29969486]
31. Saunders MJ, Wingfield T, Datta S, Montoya R, Ramos E, Baldwin MR, et al. A household-level score to predict the risk of tuberculosis among contacts of patients with tuberculosis: a derivation and external validation prospective cohort study. Lancet Infect Dis 2020 Jan;20(1):110-122. [doi: 10.1016/s1473-3099(19)30423-2]
32. Saunders MJ, Wingfield T, Tovar MA, Baldwin MR, Datta S, Zevallos K, et al. A score to predict and stratify risk of tuberculosis in adult contacts of tuberculosis index cases: a prospective derivation and external validation cohort study. Lancet Infect Dis 2017 Nov;17(11):1190-1199. [doi: 10.1016/s1473-3099(17)30447-4]
33. World Health Organization. Global Tuberculosis Report 2017. Geneva: World Health Organization; 2017.
34. World Health Organization. Global Tuberculosis Report 2018. Geneva: World Health Organization; 2018.
35. World Health Organization. Global Tuberculosis Report 2019. Geneva: World Health Organization; 2019.
36. Barron Millar E, Singhal D, Vijayaraghavan P, Seshadri S, Smith E, Dixon P, et al. Health anxiety, coping mechanisms and COVID 19: an Indian community sample at week 1 of lockdown. PLoS One 2021 Apr 21;16(4):e0250336 [FREE Full text] [doi: 10.1371/journal.pone.0250336] [Medline: 33882109]

Abbreviations

CFA: confirmatory factor analysis
CFI: cumulative fit index
CXR: chest X-ray
mHealth: mobile health
QR: quick response
RMSEA: root-mean-square error of approximation
SEM: structural equation modeling
SRMR: standardized root-mean-square residual
TB: tuberculosis
WHO: World Health Organization

Edited by A Mavragani; submitted 07.03.22; peer-reviewed by A Montilha, A Kibballi Madhukeshwar; comments to author 25.04.22; revised version received 05.05.22; accepted 05.05.22; published 07.06.22.

Please cite as:
Htet KKK, Phyu AN, Thwin T, Chongsuvivatwong V
Mobile Health App for Tuberculosis Screening and Compliance to Undergo Chest X-ray Examination Among Presumptive Cases Detected by the App in Myanmar: Usability Study
JMIR Form Res 2022;6(6):e37779
URL: https://formative.jmir.org/2022/6/e37779
doi:10.2196/37779
PMID:35623000

Original Paper

Evaluation of Dietary Management Using Artificial Intelligence and Human Interventions: Nonrandomized Controlled Trial

Fusae Okaniwa[1], PhD; Hiroshi Yoshida[2], MA

[1]Department of Theoretical Social Security Research, National Institute of Population and Social Security Research, Tokyo, Japan

[2]Graduate School of Economics and Management, Tohoku University, Miyagi, Japan

Corresponding Author:
Fusae Okaniwa, PhD
Department of Theoretical Social Security Research
National Institute of Population and Social Security Research
2-2-3 Uchisaiwaicho, Chiyoda-ku
Tokyo, 100-0011
Japan
Phone: 81 3 3595 2984
Email: okaniwa-fusae@ipss.go.jp

Abstract

Background: There has been an increase in personal health records with the increased use of wearable devices and smartphone apps to improve health. Traditional health promotion programs by human professionals have limitations in terms of cost and reach. Due to labor shortages and to save costs, there has been a growing emphasis in the medical field on building health guidance systems using artificial intelligence (AI). AI will replace advanced human tasks to some extent in the future. However, it is difficult to sustain behavioral change through technology alone at present.

Objective: This study investigates whether AI alone can effectively encourage healthy behaviors or whether human interventions are needed to achieve and sustain health-related behavioral change. We examined the effectiveness of AI and human interventions to encourage dietary management behaviors. In addition, we elucidated the conditions for maximizing the effect of AI on health improvement. We hypothesized that the combination of AI and human interventions will maximize their effectiveness.

Methods: We conducted a 3-month experiment by recruiting participants who were users of a smartphone diet management app. We recruited 102 participants and divided them into 3 groups. Treatment group I received text messages using the standard features of the app (AI-based text message intervention). Treatment group II received video messages from a companion, in addition to the text messages (combined text message and human video message intervention by AI). The control group used the app to keep a dietary record, but no feedback was provided (no intervention). We examine the participants' continuity and the effects on physical indicators.

Results: Combined AI and video messaging (treatment group II) led to a lower dropout rate from the program compared to the control group, and the Cox proportional-hazards model estimate showed a hazard ratio (HR) of 0.078, which was statistically significant at the 5% level. Further, human intervention with AI and video messaging significantly reduced the body fat percentage (BFP) of participants after 3 months compared to the control group, and the rate of reduction was greater in the group with more individualized intervention. The AI-based text messages affected the BMI but had no significant effect on the BFP.

Conclusions: This experiment shows that it is challenging to sustain participants' healthy behavior with AI intervention alone. The results also suggest that even if the health information conveyed is the same, the information conveyed by humans and AI is more effective in improving health than the information sent by AI alone. The support received from the companion in the form of video messages may have promoted voluntary health behaviors. It is noteworthy that companions were competent, even though they were nonexperts. This means that person-to-person communication is crucial for health interventions.

(JMIR Form Res 2022;6(6):e30630) doi:10.2196/30630

KEYWORDS

health promotion; dietary management; intervention; artificial intelligence; body fat percentage; body mass index; behavioral economics; nonprofessional; Japan

Introduction

Recent years have seen health promotion with the help of technological advancement, such as the use of wearable devices and smartphone apps that record an individual's health information. However, people struggle to adopt and maintain healthy behaviors [1]. Since 2016, the Japanese government has been providing a large-scale subsidy for the development of artificial intelligence (AI) in the medical field, with emphasis on the use of AI-based personal health records and the creation of health guidance systems. However, as medical and health care AI is still a nascent field, these are yet to be applied in practice and concrete examples explaining their effective use have not been made public yet [2,3].

In this study, we examine the effect of AI intervention on dietary management and elucidate the conditions for maximizing its effectiveness. Focusing on the continuity of health promotion activities and changes in physical indicators, survival analysis, and ordinary least squares (OLS) regression are conducted. We find that for advanced technology to be fully effective, it is necessary to add human intervention and customized care to AI-based interventions. If the incorporation of human interventions significantly improves physical indicators, it suggests that effective AI intervention necessarily requires unique mechanisms with added customized care. Since the COVID-19 pandemic has restricted people to their homes, the value of remote services has increased [4-6]. Online coaching, which links AI and human intervention, encourages healthy behaviors. Similarly, AI intervention could be the initiation of a new business model in this field. It is also an excellent opportunity for people with rare diseases to connect with each other globally. In this study, we focus on the next stage of the challenge in the context of enhancing the outcomes derived from the use of developed technologies. Identifying services that could be more effective with AI intervention will help to drive further technological development and stimulate new demand [7].

Our contributions to this area of study are as follows. First, this study examines the potential effectiveness of AI intervention on health promotion. Previous studies on information and communications technology (ICT) and health interventions have mainly discussed the intervention effects based on text messages [8-10] and health disparities associated with information disparities [11-14]. The main focus of our study is to develop AI-based health intervention systems. The reason for focusing on ICT and health interventions is to examine whether AI alone is sufficient to replace human intervention in this field, and to analyze the elements that can be added to maximize its effectiveness. Specifically, we examine the effectiveness of adding human interventions. Second, we examine the effectiveness of interventions through video messages. Text messages sent through short messaging service (SMS) are useful for health behaviors in the short term [8,10] and result in an increased retention of health behaviors across age and nationality [9]. However, there are not many studies on video message interventions. It should be noted that individuals with lower levels of education might not read the entire text messages [15,16], and those who are less enthusiastic about behavioral change may be more interested in video messages [17]. In previous studies, at the time of the experiments, some participants may have felt uneasy about receiving a video in place of an SMS message or a voice call due to the high cost of roaming data, which can lead to high participant attrition rates that may result in research bias [18,19]. Currently, the introduction of fifth-generation mobile communication system services in developed countries has made video viewing faster and cheaper, which may contribute toward the removal of the bias. In this sense, the effectiveness of video message interventions will increase further in the future. Third, we find that it is relevant to identify the effectiveness of the delivery of video messages by a nonprofessional through customized professional interventions. However, challenges may arise in terms of cost and reach. Exploring the possibility of nonprofessional interventions is a critical perspective when considering measures to combat labor shortages occurring due to a declining population [20]. In epidemiology and clinical fields, these issues have not been extensively studied. Moreover, these issues have not been analyzed from an economic perspective. Computerized messages are costly during development; however, once developed, they are significantly less expensive than human intervention. Moreover, the product has an extensive reach due to its availability through the internet [10,21]. Additionally, health promotion interventions are effective when their content is perceived to be personally relevant [22]. Therefore, in this study, we examine the effectiveness of nonprofessional customized messages with low-cost interventions.

In summary, the purpose of our study is to examine the effectiveness of AI and human interventions to encourage dietary management behaviors. Our study investigates whether AI alone can effectively encourage healthy behaviors or whether human interventions are needed to achieve and sustain health-related behavioral change. We elucidate the conditions for maximizing the effect of AI on health improvement. We hypothesize that the combination of AI and human interventions will maximize their effectiveness.

Methods

Standard Specifications of the App

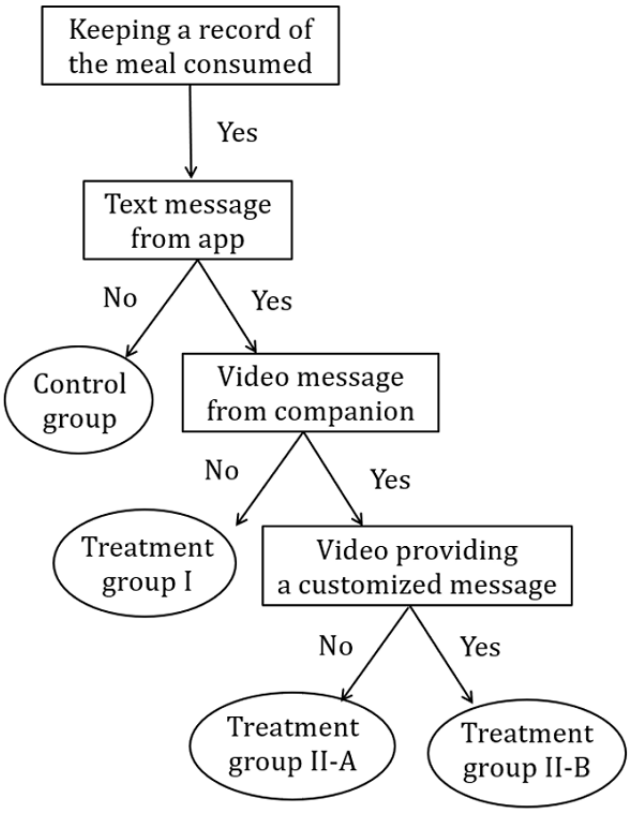
In this study, we used a diet management app for smartphones developed by asken Inc. The app functions in the following way: First, users take pictures of their meal on a smartphone. The app analyzes these pictures, registers the food automatically, and calculates the amount of nutrients and calories in the meal. Additionally, the dietary evaluation score considers ingested nutrients. The app explains its final analysis to its users with the help of graphs and advises them accordingly with a balanced diet and weight loss goals. More than 200,000 pieces of advice are developed from a proprietary algorithm based on a nutritionist's recommendations and are selected and delivered through unidirectional text messages from the app to the user. The app collects data as it is being used via (1) menus and intake of calories and nutrients, (2) dietary evaluation scores by a nutritionist in the system, (3) weight and body fat percentage (BFP) voluntarily registered by survey participants, and (4)

profiles (age, gender, and height) of participants who undertake the survey.

Interventions

When we recruited the users of the app as the participants of our study with the cooperation of asken Inc, 102 app users voluntarily participated in our experiment. They committed to record their meals on weekdays. The reason for this was to reduce the burden on participants. However, to not interfere with the behavior of motivated participants who wanted to record their holidays, we made it possible to voluntarily record their holidays. However, these holiday data were not used in this study. We promised the participants at the beginning of the experiment that we would use the weekday data. We also considered that the meals on holidays might be different from those on weekdays, but this would be reflected in the data after Monday. We conducted the experiment over a period of 3 months from February to April 2020. As shown in Figure 1, we divided the 102 app users in our study into 3 groups using the following interventions:

Figure 1. Type of intervention.



- First, in treatment group I, the standard features of the app were used. In other words, the participants recorded the dietary details of what they consumed in the app. Depending on their diet, the AI provided advice on a balanced diet from a list of 200,000 suggestions created by a confidential algorithm. For example, if a participant eats only 1 bowl of soup and records the contents in the app, it provides the following text advice: “Your meal balance score is 50. Your calorie intake is insufficient. In terms of nutritional balance, you are particularly deficient in protein and carbohydrates. For your next meal, please increase your overall food intake, focusing on meat and fish. Be sure to include 1 serving of bread or rice.” The app also provides graphs of nutrient intake and target values. These are provided via text messages.
- Second, treatment group II was provided with the same intervention as treatment group I, with the addition of a video of a human reading a text message by the app. Participants recorded their dietary details, reviewed the text advice provided by the app, and then took a screenshot of the text advice on their smartphone. They were then required to send the screenshot to their companion using the LINE app. The specifications of the app did not allow the companion advance knowledge of the text advice that would be provided to the participant. The companion was required to confirm the content of the advice on the LINE app, create a video message with an oral and visual description of the advice, and send it back to the participant. This series of events was the intervention for treatment group II. The companion is a nonprofessional, and the message is not beyond the advice of a nutritionist. They act as a human intervention to the AI text message.
- Third, we used a control group that used the app to record their meals, but no advice was delivered to them during our experiment. Therefore, participants in the control group took care of their diet and health based on their records.

Given the results of the first 2 months of the experiment, we introduced another intervention during the third month. Treatment group II was further divided into 2 subgroups: treatment group II-A received the same intervention as before, and treatment group II-B received customized care. In the former case, the companion read out a part of the text message through

AI intervention. However, in the case of the latter, the companion graphed the dietary evaluation scores and provided encouraging messages based on each person's efforts. Moreover, when participants sent us online messages expressing their impressions, the companion responded to them with a video message and delivered flexible and personalized messages.

Participants were allowed to choose which group they wanted to participate in. Treatment group I used the standard specifications of the app, meaning that there was no new burden for participants who were already using the app. Participants in treatment group II, however, were required to perform additional operations related to taking and sending screenshots, which was a new burden. We determined that it was necessary to let the participants decide for themselves whether to accept this burden.

Outcome

To understand the effects of the intervention, we analyzed 2 outcomes: (1) continuity of the program and (2) physical indicators. We used the BFP and BMI as physical indicators. In the questionnaire survey regarding the purpose of using the app, most participants responded with “to develop good eating habits” and “to lose weight.” Therefore, as outcomes, the effect of “good eating habits” was confirmed by continuity and the effect of weight loss was confirmed by the BFP and BMI [23].

We used the number of meals recorded by the participants as a measure of continuity. The experiment’s total duration was 13 weeks. We recorded their meals for more than 5 days consecutively, excluding holidays. If the participants did not record their meals for even 5 days, they were regarded as dropouts.

The BMI is a measure of nutritional status in adults. Although it is a simple indicator that can be calculated from the height and weight of an individual alone, it is not a complete physical status indicator, because it does not predict the composition of a body accurately [24,25]. However, the BFP is a more reliable health indicator than the BMI because it calculates fat as a percentage of body weight and represents its composition. The American College of Physicians has stated that the BFP is more important than the BMI in assessing a patient’s health and mortality risk [26]. Therefore, we have used the BFP as an essential health indicator in this study. Additionally, the BMI was used as a supplement to measure short-term effects. We used the average of the first week of our study as a reference point and focused on the rate of change after 3 months. As we changed the methods of treatment group II in the last month of our study, we also examined the rate of change in the average value, comparing the last week of March to the last week of April. The descriptive statistics for each variable are shown in Table S1 in Multimedia Appendix 1.

Analytical Approach

We analyzed the effects of the intervention on continuity and individuals' health indicators. First, for continuity of health behaviors, we executed a survival analysis. We regarded dropouts in our experiment as deaths, as in the case of a general survival analysis. The number of dropouts is shown in Table 1. First, we used a nonparametric model (Kaplan-Meier analysis) that did not assume a specific distribution for survival time and did not examine the effects of the covariates. We also analyzed the data with a semiparametric model (Cox proportional-hazards model [27]) that did not assume a specific distribution on survival time but estimated the parameters of the covariates and examined an effect on survival time. We created dummy variables for each group and examined the effects of each intervention. The Cox proportional-hazards model is expressed as

$$h(t|T_{ij}, X_i) = H_0(t) \exp(\beta_{ij}T_{ij} + \gamma_{ik}X_i),$$

where $h_0(t)$ is the baseline hazard; “i” denotes the participant number; “j” denotes the treatment group number; β and γ are hazard ratios (HRs); T is the treatment group dummy; and X is the set of control variables (age, gender, height).

Second, we conducted OLS regressions for health indicators:

$$R_{ik} = \alpha_i + \delta_{ij}T_{ij} + \theta_i X_i + \varepsilon_i$$

where R_k is the rate of change in the BFP (k=1) or the BMI (k=2) in experimental periods; δ and θ are coefficients vectors; and ε is an error term. We focused on the changes after 3 months. Moreover, we examined the effects of changing the intervention method for treatment group II during the last month.

Table 1. Number of dropouts and test for equality of survivor functions. Results of the estimation using a nonparametric model (Kaplan-Meier model) were as follows: Pr>χ2=0.012 (log-rank test) and 0.011 (Wilcoxon test), where Pr refers to probability.

	Control group (N=34)	Treatment group I (N=34)	Treatment group II (N=34)
Week[a], n (%)			
1	1 (3)	N/A[b]	N/A
2	1 (3)	1 (3)	N/A
3	N/A	1 (3)	N/A
4	1 (3)	N/A	N/A
5	N/A	N/A	N/A
6	N/A	N/A	N/A
7	1 (3)	2 (6)	N/A
8	1 (3)	N/A	N/A
9	N/A	2 (6)	N/A
10	2 (6)	2 (6)	N/A
11	2 (6)	N/A	N/A
12	N/A	2 (6)	1 (3)
13	N/A	N/A	N/A
Total dropouts, n (%)	9 (27)	10 (29)	1 (3)
Dropouts expected, n (%)	6.32 (19)	6.37 (19)	7.32 (22)
Sum of ranks	258	337	–595

[a]The experiment's total duration was 13 weeks (from February to April), and participants were considered to have dropped out of the program if they did not record their meals for more than 5 days consecutively, excluding holidays.

[b]N/A: not applicable.

Ethical Approval

The research ethics committee of the Graduate school of Tohoku University approved this study (approval date: January 31, 2020). We experimented with a noncontact and noninvasive approach. Recruitment was conducted by asken Inc on our behalf, and we did not have any contact with the participants. This study sought to collect and analyze information registered on the app with the permission of the participants. Necessary and sufficient information was provided to the participants, and asken Inc obtained consent from them through the app before providing the data to us. The advice communicated by the companion did not exceed the scope of the textual advice independently generated by the app designed based on a nutritionist's advice. There were no new risks inherent in this experiment, and the safety of the experiment was ensured.

Results

Effects on Continuity

Figure 2 shows the survival curve based on the Kaplan-Meier model. The vertical axis is the survival rate (here representing continuity), and the horizontal axis is the duration of the experiment (weeks). Treatment group II had the highest survival rate, followed by the control group and treatment group I. The survival curves of the 3 groups were examined for statistical significance. Results of the generalized Wilcoxon test and the log-rank test showed a significant difference between the 3 groups' survival curves at the 5% level of significance (Table 1).

Additionally, we estimated a Cox proportional-hazards model to account for the effect of the covariates (Table 2). The number of observations was 102, the number of failures was 20, and the times at risk were 1223. The HR for treatment group II was statistically significant at 0.078, that is, the dropout rate was 0.078 times higher (92% reduction) compared to the control group. This means that treatment group II had 92% fewer dropouts compared to the control group. However, for treatment group I, the HR was <1 but not statistically significant. Thus, receiving only text messages (treatment group I) does not represent a significant difference in persistence compared to receiving no intervention (control group).

Figure 2. Kaplan-Meier survival estimates. Treatment group I received a text message intervention by AI, and treatment group II received a customized video message intervention along with the text message intervention. AI: artificial intelligence.

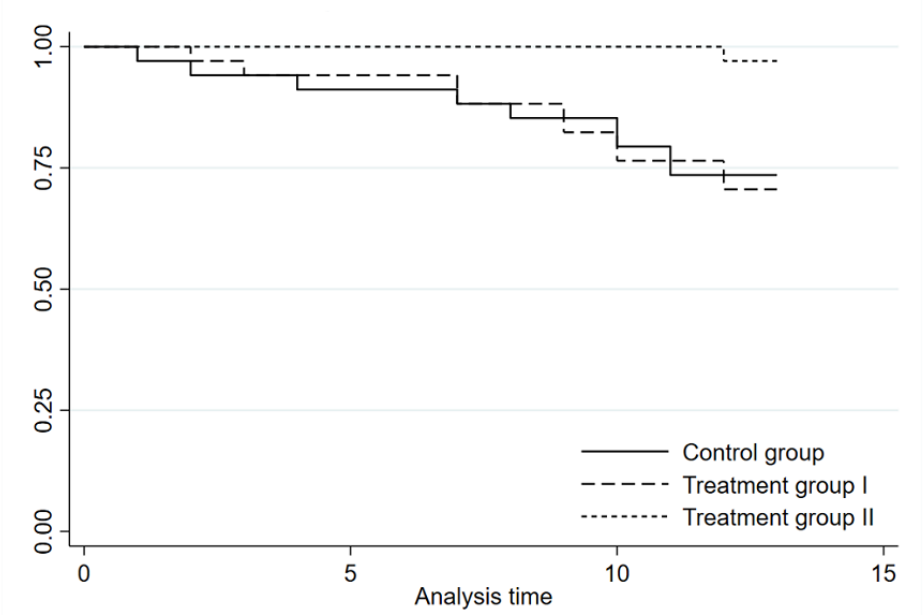


Table 2. Results of HRs[a] from the Cox proportional-hazards model (Pr>χ2=.009).

	HR	95% CI	*P* value
Treatment group I[b]	0.975	0.391	.96
Treatment group II[c]	0.078	0.009	.02
Age (years)	0.970	0.930	.17
Male	2.423	0.643	.19
Height	0.941	0.863	.17

[a]HR: hazard ratio.

[b]Treatment group I received a text message intervention by artificial intelligence.

[c]Treatment group II received a video message intervention along with the text message intervention.

Effects on Physical Indicators

After 3 months (Table 3), the BFP of treatment group II significantly decreased, but it was not statistically significant compared to treatment group I. However, the BMI was statistically significant for treatment group I but insignificant for treatment group II after 3 months. As there was no significant difference in human intervention during the first 2 months, we introduced another intervention during the third month. Treatment group II was further divided into 2 subgroups: treatment group II-A received the same intervention as before, and treatment group II-B received customized care. Observing the effects of the change in the intervention method in the last 1 month on treatment group II showed that both treatment groups II-A and II-B were statistically negatively significant for the BFP. The reduction was higher for treatment group II-B, which received a more customized intervention. However, the reduction for treatment group I was not statistically significant with AI intervention. For the BMI in the last 1 month, neither treatment group II-A nor treatment group II-B was statistically significant but treatment group I was statistically negatively significant. The number of observations during this period was as follows: BFP (3 months), 35; BMI (3 months), 64; BFP (last 1 month), 34; and BMI (last 1 month), 60. In addition, the adjusted R^2 values were 0.009, 0.072, 0.170, and 0.057, respectively.

Indicators of the effect size can be derived by several methods of analysis, such as the *t* test, ANOVA, and multiple regression analysis. In this study, multiple regression analysis was used to include more information in the estimation, and we used adjusted R^2 as an indicator of the effect size. Only the effect size of the last month's BFP was medium, but all other effect sizes were small [28]. In addition, we controlled for age and gender as personal characteristics that could affect the outcome. The physical activity level could not be included. The data we were able to collect from the app were rough, and therefore, there was little difference between participants.

Table 3. Regression results of physical indicators (OLS[a] estimation).

	After 3 months				Last 1 month			
	BFP[b]		BMI		BFP		BMI	
	Coefficient (SE)	*P* value	Coefficient (SE)	*P* value	Coefficient (SE)	*P* value	Coefficient (SE)	*P* value
Treatment group I[c]	–0.029 (0.056)	.61	–0.026 (0.010)	.44	–0.013 (0.045)	.77	–0.009 (0.004)	.04
Treatment group II[d]	–0.102 (0.050)	.05	–0.008 (0.010)	.01	N/A	N/A	N/A	N/A
Treatment group II-A[e]	N/A[f]	N/A	N/A	N/A	–0.082 (0.044)	.07	–0.003 (0.005)	.50
Treatment group II-B[g]	N/A	N/A	N/A	N/A	–0.218 (0.074)	.01	0.007 (0.009)	.43
Male	0.031 (0.048)	.52	–0.011 (0.009)	.22	0.028 (0.041)	.49	–0.007 (0.004)	.10
Age (years)	–0.001 (0.002)	.59	0.000 (0.000)	.31	–0.001 (0.002)	.78	0.000 (0.000)	.75
Constant	0.019 (0.110)	.86	0.013 (0.017)	.46	0.027 (0.090)	.77	0.006 (0.008)	.51

[a]OLS: ordinary least squares.

[b]BFP: body fat percentage.

[c]Treatment group I received a text message intervention by artificial intelligence.

[d]Treatment group II received a video message intervention along with the text message intervention. During the last month, treatment group II was divided into 2 subgroups.

[e]Treatment group II-A received a video intervention that only read out text messages as before.

[f]N/A: not applicable.

[g]Treatment group II-B received a more customized intervention.

Discussion

Principal Findings

A society in which individuals use a variety of new devices to engage in health promotion activities in their daily lives was envisioned in this study. To effectively implement advanced technologies in health promotion, it is important to know what interventions might encourage people to engage in autonomous health activities. We examined the effectiveness of interventions by AI and humans to encourage healthy behaviors. Our study compared continuity of the program and the effects on health indicators using an AI-based intervention, with text messages and customized (nonexpert) video messages added to the intervention for the users of a dietary management app. We found that combined AI and video messaging (treatment group II) led to a lower dropout rate from the program compared to the control group, and the Cox proportional-hazards model estimate showed an HR of 0.078, which was statistically significant at the 5% level. Further, human intervention with AI and video messaging significantly reduced the BFP of participants after 3 months compared to the control group, and the rate of reduction was greater in the group with more individualized intervention. The AI-based text messages affected the BMI but had no significant effect on the BFP.

Considerations and Future Directions

First, this study shows that it is challenging to sustain healthy behaviors with AI intervention alone. We also found that health improvement is the highest when the intervention is delivered with human video messages and AI text messages. Traditional economic theory assumes that people behave rationally after receiving accurate information. Behavioral economics, however, calls this rational world bias and considers that, in reality, it is challenging to take a rational action [29-31]. The results of this study suggest the existence of a rational world bias in ongoing healthy behavior and long-term health improvement; providing information through AI alone may only have short-term effects.

Second, adding human intervention to AI intervention by delivering videos of an individual reading AI-based text messages (treatment group II-A) had a significantly positive impact on persistence and health promotion effects. This means that even if the information conveyed is the same, differences in the delivery method produces differences in effectiveness. Here, human-communicated information was more effective in promoting healthy behavior.

Third, the effect was higher when the companion increased the level of customization, such as by mentioning the participants' names in the videos while cheering them on and plotting the trends in a graph in their respective dietary evaluation scores (treatment group II-B). This can be considered a type of coaching effect. Highly individualized interventions are effective [32], and direct coaching is also effective [33]. Further, this study confirmed the effectiveness of remote coaching.

Fourth, it is also important to note that the companions were capable, even though they were not professionals in the field. This means that the process of human communication is as important as the content of the information. Due to the effectiveness of nonprofessional interventions, there is potential to address labor shortages due to a decline in the future population and human costs.

Fifth, behavioral economics believes that humans are limitedly rational beings and tend to have base rate neglect [34]. For example, even if obesity is known to increase the risk of acquiring diseases in the future, it is difficult to continue health behaviors. This may be especially true for people with high

time preference, that is, individuals who prioritize present utility over future utility. In this experiment, nonexperts supporting users every day increased the continuity effect. For those with high time preference rates, the daily support received from their companions may have been a reward and an incentive to continue their long-term health behaviors. It may be useful to consider effective interventions for people with high time preference rates who have difficulty sustaining healthy behaviors.

Finally, we also infer that each intervention has different effects on the BFP and BMI. The BMI is calculated based on the height and weight of individuals only, and it cannot distinguish between fat and muscle. Therefore, it is an indicator that quantifies the apparent body size. The height of an adult does not change immediately, but the weight fluctuates with the content and frequency of meals per day, which can result in an increase or decrease in the BMI. However, the BFP is a measure of body composition; it cannot be reduced immediately and requires continuous effort. This difference is evident from the fact that the correlation between BMI and BFP is not always strong [35]. The AI text messages had a significant and negative effect on the BMI but no statistically significant effect on the BFP. This suggests that text messages may have effectively promoted weight loss in the short term, but thi may not hold true in the long term. On the contrary, customized video messages did not affect the BMI and had a significantly negative effect on the BFP, which implies that the video messages were effective in promoting ongoing health behaviors.

Limitations

There are several limitations to the interpretation of our results. First, there exists a sampling bias. We recruited app users as participants and also asked them to choose the intervention group they wanted to belong to. These aspects may have affected our results. Second, participants may have been highly conscious about their health; hence, the results of our study may be valid only for those who are already health conscious. The effect on people with low health consciousness requires further study. Owing to limitations in the data we collected, we were not able to include in our estimates many confounding variables that could affect outcomes, and the effect size was also small. Approximately 100 users participated in our study, which is not a sufficient size; hence, it may be difficult to generalize the results obtained.

Second, since weight and BFP data were obtained voluntarily, only highly voluntary and health-conscious individuals may have recorded the data. On the last day of the experiment, 60 (58.8%) of the 102 participants continued recording their weights and 34 (33.3%) continued recording their BFP. However, dietary records continued at 82 (80.4%) of the 102 participants in the last month. Therefore, bias may have occurred due to voluntary collection of data. AI and human interventions involved a procedure in which participants were required to send a screenshot of a text message to the companion. The procedure was carried out because our experiment was conducted without changing the specifications of the app. This may have increased the spontaneity of the participants.

Third, we should note the validity of the outcome. The app used in the experiment and the advice provided were designed to shape a balanced diet. Therefore, it is plausible to use an indicator that is related to diet quality. The app delivers the dietary evaluation scores calculated by the proprietary algorithm of asken Inc via text messages. However, we were unable to confirm the objective validity of this score. The results of our survey showed that the participants were interested in healthy living in a broader sense and not by merely being on a balanced diet. Therefore, we used objective health indicators as outcomes. The collection of adequate objective data on dietary quality is a challenge for future research.

Finally, the analysis did not include the costs of human interventions. In our study, the combination of AI and human interventions had the highest effect. However, human intervention is costly. No matter how significant the effect is, the behavioral change may not necessarily sustain if users and companions pay high costs (eg, in terms of money, time, human resources, and efforts) to take advantage of it. Therefore, it is necessary to examine human intervention in conjunction with the costs.

Conclusion

This experiment shows that it is challenging to sustain participants' healthy behavior with AI intervention alone. The results also suggest that even if the health information conveyed is the same, the information conveyed by humans is more effective in improving health than the information sent by AI. The support received from the companion in the form of video messages may have promoted voluntary health behaviors. It is noteworthy that the companions were competent, even though they were nonexperts. This means that the process of person-to-person communication is crucial for health interventions. The results of this experiment show both short- and long-term effects, which may help us consider effective intervention strategies that respond to these differences with respect to time preference.

Acknowledgments

We express our gratitude to all the participants of our experiment. We would like to thank Professor Shuzo Nishimura of the Kyoto University of Advanced Science and Associate Professor Shusaku Sasaki of Tohoku Gakuin University for their helpful contributions at the Japan Institute of Public Finance and the Japan Health Economics Association. We would also like to thank Associate Professor Michio Yuda and Assistant Professor Chen Fengming of Tohoku University for their helpful comments.

This work was supported by the Japan Science and Technology Agency (JST), Center of Innovation (COI; Grant JPMJCE1303).

Data Availability

The data sets generated and analyzed in this study are not publicly available due to confidentiality and ethics.

Authors' Contributions

FO and HY worked on the study design, experiment, and final approval. FO was responsible for data analysis and drafting of the study. HY worked on the experimental methods and critical revision.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Table S1. Description of variables.
[DOCX File , 16 KB - formative_v6i6e30630_app1.docx]

References

1. Glanz K, Rimer BK, Viswanath K, editors. Health Behavior: Theory, Research, and Practice 5th Edition. California: Jossey-Bass; Jul 27, 2015.
2. Ishii E, Ebner DK, Kimura S, Agha-Mir-Salim L, Uchimido R, Celi LA. The advent of medical artificial intelligence: lessons from the Japanese approach. J Intensive Care 2020 May 18;8(1):35 [FREE Full text] [doi: 10.1186/s40560-020-00452-5] [Medline: 32467762]
3. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat Med 2019 Jan 7;25(1):30-36 [FREE Full text] [doi: 10.1038/s41591-018-0307-0] [Medline: 30617336]
4. Perlman A, Vodonos Zilberg A, Bak P, Dreyfuss M, Leventer-Roberts M, Vurembrand Y, et al. Characteristics and symptoms of app users seeking COVID-19-related digital health information and remote services: retrospective cohort study. J Med Internet Res 2020 Oct 20;22(10):e23197 [FREE Full text] [doi: 10.2196/23197] [Medline: 32961527]
5. Shankar K. The impact of COVID‐19 on IT services industry: expected transformations. Brit J Manag 2020 Jul 09;31(3):450-452. [doi: 10.1111/1467-8551.12423]
6. Chamola V, Hassija V, Gupta V, Guizani M. A comprehensive review of the COVID-19 pandemic and the role of IoT, drones, AI, blockchain, and 5G in managing its impact. IEEE Access 2020 May 13;8:90225-90265. [doi: 10.1109/access.2020.2992341]
7. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J 2019 Jun 13;6(2):94-98 [FREE Full text] [doi: 10.7861/futurehosp.6-2-94] [Medline: 31363513]
8. Fjeldsoe BS, Marshall AL, Miller YD. Behavior change interventions delivered by mobile telephone short-message service. Am J Prev Med 2009 Feb;36(2):165-173. [doi: 10.1016/j.amepre.2008.09.040] [Medline: 19135907]
9. Cole-Lewis H, Kershaw T. Text messaging as a tool for behavior change in disease prevention and management. Epidemiol Rev 2010 Mar 30;32(1):56-69 [FREE Full text] [doi: 10.1093/epirev/mxq004] [Medline: 20354039]
10. Head KJ, Noar SM, Iannarino NT, Grant Harrington N. Efficacy of text messaging-based interventions for health promotion: a meta-analysis. Soc Sci Med 2013 Nov;97:41-48. [doi: 10.1016/j.socscimed.2013.08.003] [Medline: 24161087]
11. Viswanath K. Cyberinfrastructure: an extraordinary opportunity to bridge health and communication inequalities? Am J Prev Med 2011 May;40(5 Suppl 2):S245-S248. [doi: 10.1016/j.amepre.2011.02.005] [Medline: 21521602]
12. Ackerson LK, Ramanadhan S, Arya M, Viswanath K. Social disparities, communication inequalities, and HIV/AIDS-related knowledge and attitudes in India. AIDS Behav 2012 Oct 26;16(7):2072-2081. [doi: 10.1007/s10461-011-0031-y] [Medline: 21870161]
13. Jung M, Arya M, Viswanath K. Effect of media use on HIV/AIDS-related knowledge and condom use in sub-Saharan Africa: a cross-sectional study. PLoS One 2013 Jul 12;8(7):e68359 [FREE Full text] [doi: 10.1371/journal.pone.0068359] [Medline: 23874598]
14. Niederdeppe J, Bigman CA, Gonzales AL, Gollust SE. Communication about health disparities in the mass media. J Commun 2013 Jan 10;63(1):8-30. [doi: 10.1111/jcom.12003]
15. Liu Z. Reading behavior in the digital environment: changes in reading behavior over the past ten years. J Doc 2005;61(6):700-712. [doi: 10.1108/00220410510632040]
16. Vandelanotte C, Mummery WK. Qualitative and quantitative research into the development and feasibility of a video-tailored physical activity intervention. Int J Behav Nutr Phys Act 2011 Jul 01;8(1):70 [FREE Full text] [doi: 10.1186/1479-5868-8-70] [Medline: 21718546]
17. Stanczyk N, Bolman C, van Adrichem M, Candel M, Muris J, de Vries H. Comparison of text and video computer-tailored interventions for smoking cessation: randomized controlled trial. J Med Internet Res 2014 Mar 03;16(3):e69 [FREE Full text] [doi: 10.2196/jmir.3016] [Medline: 24589938]

18. Danaher BG, Jazdzewski SA, McKay HG, Hudson CR. Bandwidth constraints to using video and other rich media in behavior change websites. J Med Internet Res 2005 Sep 16;7(4):e49 [FREE Full text] [doi: 10.2196/jmir.7.4.e49] [Medline: 16236701]
19. Whittaker R, Dorey E, Bramley D, Bullen C, Denny S, Elley CR, et al. A theory-based video messaging mobile phone intervention for smoking cessation: randomized controlled trial. J Med Internet Res 2011 Jan 21;13(1):e10 [FREE Full text] [doi: 10.2196/jmir.1553] [Medline: 21371991]
20. Johnson Z, Howell F, Molloy B. Community mothers' programme: randomised controlled trial of non-professional intervention in parenting. BMJ 1993 May 29;306(6890):1449-1452 [FREE Full text] [doi: 10.1136/bmj.306.6890.1449] [Medline: 8518642]
21. Kreuter MW, Strecher VJ, Glassman B. One size does not fit all: the case for tailoring print materials. Ann Behav Med 1999 Dec;21(4):276-283. [doi: 10.1007/bf02895958]
22. Petty R, Cacioppo J. Attitudes and Persuasion: Classic and Contemporary Approaches. New York, NY: Routledge; 1981:1-314.
23. Andreyeva T, Long MW, Henderson KE, Grode GM. Trying to lose weight: diet strategies among Americans with overweight or obesity in 1996 and 2003. J Am Diet Assoc 2010 Apr;110(4):535-542. [doi: 10.1016/j.jada.2009.12.029] [Medline: 20338279]
24. WHO Expert Consultation. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. Lancet 2004 Jan;363(9403):157-163. [doi: 10.1016/s0140-6736(03)15268-3]
25. Moreno LA, Blay MG, Rodríguez G, Blay VA, Mesana MI, Olivares JL, AVENA-Zaragoza Study Group. Screening performances of the International Obesity Task Force body mass index cut-off values in adolescents. J Am Coll Nutr 2006 Oct;25(5):403-408. [doi: 10.1080/07315724.2006.10719552] [Medline: 17031009]
26. Padwal R, Leslie WD, Lix LM, Majumdar SR. Relationship among body fat percentage, body mass index, and all-cause mortality. Ann Intern Med 2016 Mar 08;164(8):532. [doi: 10.7326/m15-1181]
27. Cox DR. Regression models and life-tables. J R Stat Soc Series B Methodol 2018 Dec 05;34(2):187-202. [doi: 10.1111/j.2517-6161.1972.tb00899.x]
28. Cohen J. Statistical Power Analysis for the Behavioral Sciences (2nd edition). New York, NY: Routeledge; 1988.
29. Simon H. Models of Bounded Rationality, Volume 2. Cambridge, MA: MIT Press; 1982:2.
30. Frank RG, Lamiraud K. Choice, price competition and complexity in markets for health insurance. J Econ Behav Organ 2009 Aug;71(2):550-562. [doi: 10.1016/j.jebo.2009.04.005]
31. Loewenstein G, Asch DA, Volpp KG. Behavioral economics holds potential to deliver better results for patients, insurers, and employers. Health Aff (Millwood) 2013 Jul;32(7):1244-1250. [doi: 10.1377/hlthaff.2012.1163] [Medline: 23836740]
32. Rawl SM, Skinner CS, Perkins SM, Springston J, Wang H, Russell KM, et al. Computer-delivered tailored intervention improves colon cancer screening knowledge and health beliefs of African-Americans. Health Educ Res 2012 Oct 27;27(5):868-885 [FREE Full text] [doi: 10.1093/her/cys094] [Medline: 22926008]
33. Boysen-Rotelli S. Coaching effectiveness: coach and coachee characteristics that lead to success. PoC 2018 Nov 30;3(2):6-26. [doi: 10.22316/poc/03.2.02]
34. Kahneman D. Thinking, Fast and Slow. New York, NY: Farrar, Straus and Giroux; 2011.
35. Meeuwsen S, Horgan G, Elia M. The relationship between BMI and percent body fat, measured by bioelectrical impedance, in a large adult sample is curvilinear and influenced by age and sex. Clin Nutr 2010 Oct;29(5):560-566. [doi: 10.1016/j.clnu.2009.12.011] [Medline: 20359792]

Abbreviations

AI: artificial intelligence
BFP: body fat percentage
HR: hazard ratio
ICT: information and communications technology
OLS: ordinary least squares
SMS: short messaging service

Edited by T Leung; submitted 26.05.21; peer-reviewed by R Kretschmann, S Lin; comments to author 14.07.21; revised version received 07.09.21; accepted 17.03.22; published 08.06.22.

Please cite as:
Okaniwa F, Yoshida H
Evaluation of Dietary Management Using Artificial Intelligence and Human Interventions: Nonrandomized Controlled Trial
JMIR Form Res 2022;6(6):e30630
URL: https://formative.jmir.org/2022/6/e30630
doi:10.2196/30630
PMID:35675107

Original Paper

Sociodemographic Characteristics Associated With an eHealth System Designed to Reduce Depressive Symptoms Among Patients With Breast or Prostate Cancer: Prospective Study

Nuhamin Gebrewold Petros[1], BA, MMSc; Gergo Hadlaczky[1], PhD; Sara Carletto[2], PhD; Sergio Gonzalez Martinez[3], MSc; Luca Ostacoli[4], PhD; Manuel Ottaviano[3], PhD; Björn Meyer[5], PhD; Enzo Pasquale Scilingo[6], PhD; Vladimir Carli[1], MD, PhD

[1]National Centre for Suicide Research and Prevention of Mental Ill-Health, Department of Learning, Informatics, Ethics and Management, Karolinska Institute, Stockholm, Sweden

[2]Department of Neuroscience "Rita Levi Montalcini", Università degli Studi di Torino, Turin, Italy

[3]Life Supporting Technologies, Universidad Politecnica de Madrid, Madrid, Spain

[4]Department of Clinical and Biological Sciences, Universita degli studi di Torino, Turin, Italy

[5]GAIA AG, Hamburg, Germany

[6]Research Center, School of Engineering, University of Pisa, Pisa, Italy

Corresponding Author:
Nuhamin Gebrewold Petros, BA, MMSc
National Centre for Suicide Research and Prevention of Mental Ill-Health
Department of Learning, Informatics, Ethics and Management
Karolinska Institute
Granits väg 4
Stockholm, 171 65
Sweden
Phone: 46 707313324
Email: nuhamin.petros@ki.se

Abstract

Background: eHealth interventions have become a topic of interest in the field of mental health owing to their increased coordination and integration of different elements of care, in treating and preventing mental ill health in patients with somatic illnesses. However, poor usability, learnability, and user engagement might affect the effectiveness of an eHealth intervention. Identifying different sociodemographic characteristics that might be associated with higher perceived usability can help improve the usability of eHealth interventions.

Objective: This study aimed to identify the sociodemographic characteristics that might be associated with the perceived usability of the NEVERMIND (Neurobehavioural Predictive and Personalised Modelling of Depressive Symptoms During Primary Somatic Diseases) eHealth system, comprising a mobile app and a sensorized shirt, in reducing comorbid depressive symptoms in patients with breast or prostate cancer.

Methods: The study included a total of 129 patients diagnosed with breast (n=80, 62%) or prostate (n=49, 38%) cancer, who received a fully automated mobile app and sensorized shirt (NEVERMIND system). Sociodemographic data on age, sex, marital status, education level, and employment status were collected at baseline. Usability outcomes included the System Usability Scale (SUS), a subjective measure that covers different aspects of system usability; the user version of the Mobile App Rating Scale (uMARS), a user experience questionnaire; and a usage index, an indicator calculated from the number of days patients used the NEVERMIND system during the study period.

Results: The analysis was based on 108 patients (n=68, 63%, patients with breast cancer and n=40, 37%, patients with prostate cancer) who used the NEVERMIND system for an average of 12 weeks and completed the study. The overall mean SUS score at 12 weeks was 73.4 (SD 12.5), which indicates that the NEVERMIND system has good usability, with no statistical differences among different sociodemographic characteristics. The global uMARS score was 3.8 (SD 0.3), and women rated the app higher than men (β=.16; *P*=.03, 95% CI 0.02-0.3), after adjusting for other covariates. No other sociodemographic characteristics were

associated with higher uMARS scores. There was a statistical difference in the use of the NEVERMIND system between women and men. Women had significantly lower use (β=–0.13; *P*=.04, 95% CI −0.25 to −0.01), after adjusting for other covariates.

Conclusions: The findings suggest that the NEVERMIND system has good usability according to the SUS and uMARS scores. There was a higher favorability of mobile apps among women than among men. However, men had significantly higher use of the NEVERMIND system. Despite the small sample size and low variability, there is an indication that the NEVERMIND system does not suffer from the *digital divide*, where certain sociodemographic characteristics are more associated with higher usability.

Trial Registration: German Clinical Trials Register RKS00013391; https://www.drks.de/drks_web/navigate.do?navigationId=trial.HTML&TRIAL_ID=DRKS00013391

(JMIR Form Res 2022;6(6):e33734) doi:10.2196/33734

KEYWORDS

mental health; depression; eHealth; usability; breast cancer; prostate cancer; System Usability Scale; SUS; the user version of the Mobile App Rating Scale; uMARS; Neurobehavioural Predictive and Personalised Modelling of Depressive Symptoms During Primary Somatic Diseases; NEVERMIND system

Introduction

Background

Physical illnesses such as cancer take a toll on a patient's physical well-being and mental health [1]. The 1-year prevalence of depression is approximately 3% in the general population and between 8% and 24% in patients diagnosed with cancer [2]. Mental ill health, especially depressive symptoms, can affect the quality of life and response to treatment and prognosis in patients diagnosed with breast or prostate cancer, subsequently affecting the prognosis of the cancer outcome [3,4]. Consequently, eHealth and information and communication technology–based self-management tools are of interest to the field of mental health because of their increased engagement with patients, faster response times, and increased coordination and integration of different elements of care [5-7]. These eHealth interventions are designed to curb mental ill health–related consequences at the individual, societal, and health care levels [8-10]. Understanding different aspects of the usability of eHealth interventions could provide substantial clinical benefits [11]. For example, if the technology has poor usability and learnability and low user engagement, the overall effectiveness on patient outcomes may be low, even if the clinical content of the intervention is otherwise effective [12]. Identifying barriers to and facilitators of the implementation process has the potential to streamline eHealth interventions to deliver the intended clinical content optimally. The International Organization for Standardization, an organization that measures and certifies the quality, safety, and efficiency of products, services, and systems, defines usability as "the extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" [13]. To determine the usability of any new eHealth technology, rigorously developed and appropriate measures must be chosen [12]. One method of determining usability is to identify the different sociodemographic variables associated with better or higher use of the eHealth intervention, measured using appropriate questionnaires and use data. A recent literature review showed that older age, lower income, lower education, living alone, and living in rural areas were associated with lower eHealth intervention use in patients diagnosed with chronic disorders [14]. It is advantageous to investigate the different sociodemographic characteristics that can specifically influence the use of newly developed eHealth interventions.

Objectives

The objective of this study was to determine the different sociodemographic characteristics that are associated with the perceived usability of the NEVERMIND (Neurobehavioural Predictive and Personalised Modelling of Depressive Symptoms During Primary Somatic Diseases) system, a newly developed eHealth system for helping patients diagnosed with kidney failure, myocardial infarction, leg amputation, and breast and prostate cancer to self-manage their mental health symptoms, including depressive symptoms. As the NEVERMIND system is a newly developed system, the usability and effectiveness of the system need to be investigated. The aim of our study was to identify different sociodemographic variables that might be associated with the perceived usability of the NEVERMIND system in patients with breast or prostate cancer.

Methods

Overview

This study used data from the European Union–funded Horizon 2020 project, NEVERMIND. NEVERMIND uses information and communication technology–enabled self-management procedures. The NEVERMIND system comprises a sensorized shirt to collect biomedical data (electrocardiogram, respiration dynamics, and body movement), and a user interface in the form of a mobile app to collect data on mental health symptoms (depressive and anxiety-related symptoms, stress, and sleep problems) using mood-assessing psychometric questionnaires (Figure 1). Data from the questionnaires and biomedical data are used to predict patients' depressive symptoms, to provide effective feedback and recommendations (Figure 2). This feedback includes personalized lifestyle behavioral feedback on physical activity, sleep hygiene, dietary habits, mindfulness practice, and cognitive behavior therapy training.

The effectiveness of the NEVERMIND system was evaluated in a randomized controlled trial of 425 patients aged ≥18 years. Patients diagnosed with breast cancer, prostate cancer, myocardial infarction, kidney failure, or leg amputation were

recruited from clinical centers in Turin and Pisa, Italy, and Lisbon, Portugal, from November 2017 to December 2019. The results of the randomized controlled trial showed that the NEVERMIND system was superior to standard care in reducing depressive symptoms in patients diagnosed with severe somatic illnesses [15]. In the randomized controlled trial, patients were allocated to either receive the NEVERMIND system in addition to standard care or receive standard care only. Patients in the NEVERMIND intervention received a mobile phone with the NEVERMIND app on it and the sensorized shirt at the recruitment center. The mobile phone that the patients received was configured to use only the app. Patients also completed baseline sociodemographic and usability questionnaires. The usability, acceptability, and satisfaction questionnaires were administered at 2 time points in the study to evaluate specific aspects and assess the subjective quality of the NEVERMIND system. Patients in the NEVERMIND intervention group used the NEVERMIND system for 12 weeks; usability questionnaires were administered at 4 weeks and at the end of 12 weeks.

Figure 1. Welcome page of the NEVERMIND (Neurobehavioural Predictive and Personalised Modelling of Depressive Symptoms During Primary Somatic Diseases) mobile app.

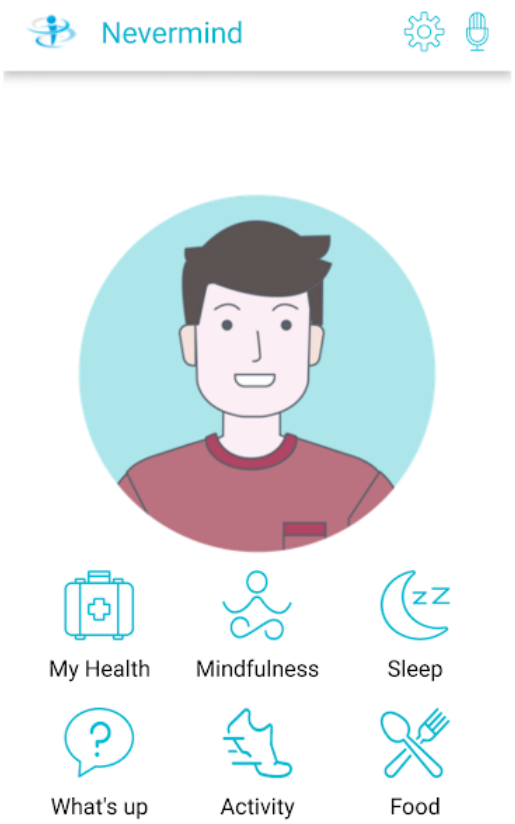
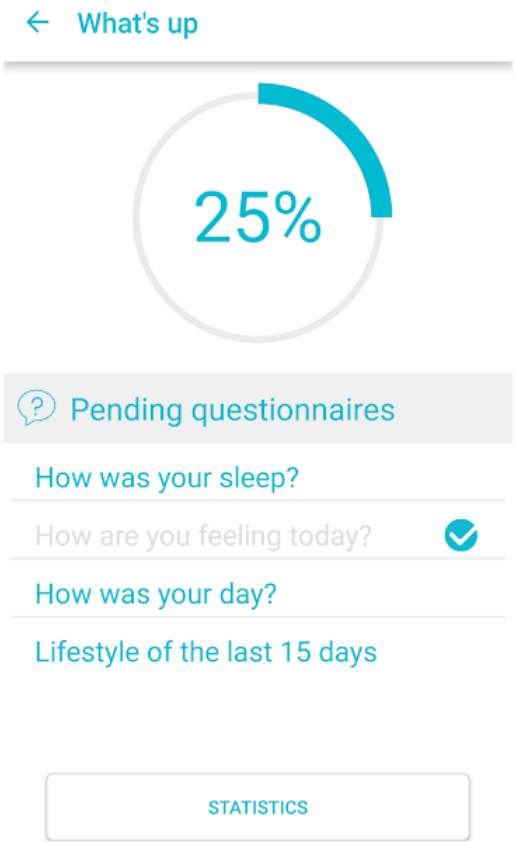


Figure 2. An example of questions administered on the NEVERMIND (Neurobehavioural Predictive and Personalised Modelling of Depressive Symptoms During Primary Somatic Diseases) mobile app.



Recruitment

The inclusion and exclusion criteria for the NEVERMIND trial are included in the published protocol [10]. The following inclusion and exclusion criteria refer to the selection of patients for this study. In this study, patients with breast or prostate cancer were recruited from the Piedmont Oncological Network at San Luigi Gonzaga University Hospital, Turin, Italy, and the Breast Unit-Oncology Department and Urology Department at Città della Salute e della Scienza University Hospital, Turin, Italy.

Inclusion Criteria

Patients who were allocated to the NEVERMIND intervention group, had a diagnosis of either breast or prostate cancer, and completed the trial were included in the study.

Exclusion Criteria

Patients were excluded if they were allocated to the control group in the NEVERMIND study or if they were in the NEVERMIND intervention group but were diagnosed with other severe somatic conditions, such as kidney failure, leg amputation, and myocardial infarction, as the scope of the study was limited to patients diagnosed with cancer. Patients who belonged to the NEVERMIND intervention group but dropped out of the study before receiving the NEVERMIND system were also excluded.

Data Collection

Exposure Variables

Sociodemographic information, including age, sex, marital status, education level, employment status, and living arrangements, was recorded at baseline.

Outcome

Overview

Three usability metrics were used as the outcome measures to evaluate the usability of the NEVERMIND system:

1. System Usability Scale (SUS): patients completed the SUS at 4 weeks (interim) and 12 weeks (final) after using the NEVERMIND system. The SUS is one of the most frequently used usability measurements that covers the attributes *learnability* and *satisfaction* of the usability dimensions [12,16,17]. The scale is a 10-item subjective measure that can quantify how well users have interacted with and used the product, covering the ease of use of different functionalities, and assessing any technical issues during use, the user's impression and benefits of using the system. Each item's score ranges from 0 to 4, and the sum of the items is multiplied by 2.5 to give a transformed composite scale that ranges from 0 to 100; a score of 68 is considered above average [17]. The scale has an interitem correlation of 0.34 to 0.69 and high reliability (Cronbach α=.91) [12]. The SUS is used to assess the usability of eHealth tools in different fields [18,19].

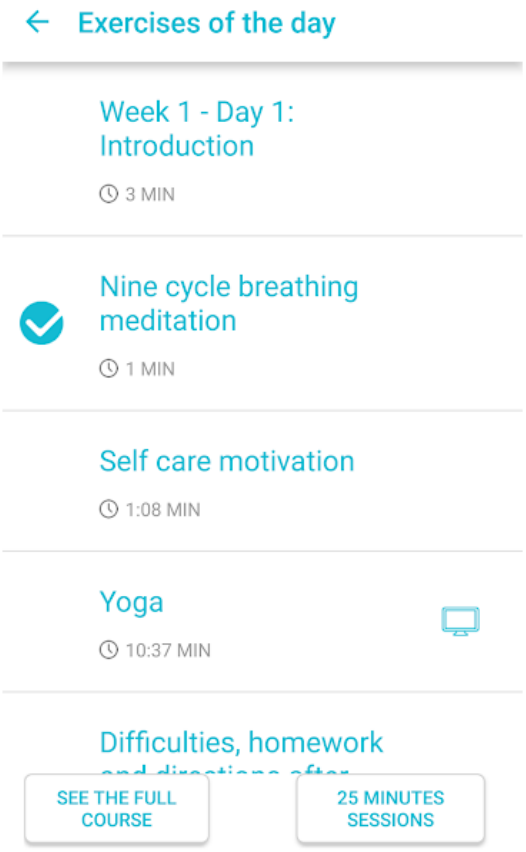
2. The user version of the Mobile App Rating Scale (uMARS): patients completed uMARS at 12 weeks after using the NEVERMIND system. uMARS is the adapted end-user version of the Mobile App Rating Scale, a scale for digital health experts that measures how good a mobile health app is in different dimensions. uMARS measures the app quality by measuring engagement, functionality, aesthetics, and information, to design and develop high-quality mobile apps [20]. The uMARS global score, as well as the 4 objective quality scales, ranges from 0 to 5, with 5 indicating the app to be of very high quality [20]. The uMARS has also been shown to have high internal consistency (Cronbach α=.90) and a good interrater reliability correlation coefficient (0.66-0.70) [20].
3. Use of the NEVERMIND system: the NEVERMIND system uses biomedical data and mental health symptoms gathered from the sensorized shirt and mobile app, respectively, to deliver personalized and timely lifestyle and behavioral feedback as well as mindfulness and cognitive behavioral therapy training in the form of different modules within the mobile app.

The 4 modules are described in the following sections.

Physical Activity

The physical activity module was designed to reinforce motivation and help the patient achieve goals established at enrollment, based on the patient's physical functionality and capacity as evaluated by a clinician. Patients had access to a list of suggested exercises in a video format, a list of previously performed exercises, and tips and recommendations guiding them on how to perform the suggested exercises (Figure 3).

Figure 3. Physical exercise module of the NEVERMIND (Neurobehavioural Predictive and Personalised Modelling of Depressive Symptoms During Primary Somatic Diseases) mobile app.



Dietary Recommendations

Similar to the physical activity module, the dietary module was designed to reinforce motivation and help patients achieve incremental goals. A clinician set these goals by considering the type of diet the patient was following, as well as the dietary preferences of the patient. Patients were recommended types of breakfast they should have and how much protein, carbohydrate, and fat they should consume, among other things. These dietary recommendations were presented in recipe videos and educational reading content.

Sleep Hygiene Practice

Patients were instructed to use a sleep agenda and report on parameters related to sleep quality. Upon opening the sleep module, patients were asked about their sleep quality during the previous night (eg, hours in bed and time to fall asleep). Patients were then directed to 4 options: going to bed, daily recommendations on sleep practice, results, and tips. Sleep practice was delivered in video or audio format. Patients were also prompted to wear the sensorized shirt while performing the sleep practice.

Mindfulness Practice

The mindfulness module included different types of mindfulness practice of different lengths. The practices offered to patients were personalized according to their disease, preferences set during their enrollment, and mental health symptoms reported in their daily questionnaires. The module also had the option of wearing the sensorized shirt during any of the mindfulness practice sessions. When the user was wearing the shirt, the app received biofeedback consisting of the user's respiratory and heart rates. This information was then displayed on the screen when the user performed the exercise. The daily recommended practice showed users one practice a day that they should try to complete.

Overall, all the modules were designed to reinforce motivation, help achieve the intended goal and in turn, improve the patient's mental health, including depressive symptoms. Each module had use data recorded by distinct days of use (the number of days a patient has used the specific module), log data (when a patient opens the app but does not necessarily use the app or the modules or sends any data to the server), and the number of completed practices (the number of completed practices within each module). A remote server also collected data from the sensorized shirt, where use data were expressed in terms of distinct days of use and log data.

Our usage index was computed by adding the number of days patients used only the app or only the shirt and days that patients used both the app and the shirt. However, as patients had different study periods, we divided the index by the number of days patients were included in the study.

Statistical Analysis

The sample size was based on the number of patients with breast or prostate cancer in the NEVERMIND study who received the NEVERMIND system. All outcomes were measured on a continuous scale, and sociodemographic characteristics were dichotomized. The variable, living arrangement, was categorized as either living alone or being a cohabitant (living with a partner, other family members, or with other people), whereas education was dichotomized as low (no education and primary or secondary school) and high (college level and above). Marital status was also dichotomized as single (unattached, divorced, separated, and widowed) and living with someone (marriage or domestic partnership), whereas employment status was categorized as either unemployed (retired, unemployed, or not working owing to other reasons, including ill health) or employed. The normality of the outcomes was checked using the Skewness and Kurtosis tests. The association between sociodemographic characteristics and SUS, uMARS, and usage index was measured using multivariate regression. All analyses were performed using STATA/MP (version 15.1; StataCorp LLC).

Ethics Approval

Ethical approval for the NEVERMIND study was submitted and approved by each of the local research ethics committees in the centers where the intervention was implemented (Pisa Comitato Etico di Area Vasta Nord Ovest (Comitato Etico Sperimentazione Farmaco; 912/2015); Ethical Committee of Città della Salute e della Scienza di Torino University Hospital and Ethical Committee of San Luigi Gonzaga University Hospital, Orbassano (185/2015); Ethics Committee of the Medical Academic Centre of the University of Lisbon (223/16). Additional ethical approval for the analysis of the pseudoanonymized data was obtained by the Swedish Ethics Review Authority (Etikprövningsmyndigheten; Dnr 2020-04175).

Results

Overview

A total of 129 patients diagnosed with either breast or prostate cancer were included in the intervention group. Of the 129 patients, 108 (83.7%) completed the study and 21 (16.3%) dropped out after the baseline assessment. Of these 21 patients, 11 (52%) dropped out before receiving the NEVERMIND system because of nickel allergy (1/21, 5%), pacemaker (1/21, 5%), emergency surgery (2/21, 10%) and not coming back to get the system (7/21, 33%). Of the remaining 21 patients, 9 (43%) received the system but did not open the mobile app or use the shirt and 1 (1/21, 5%) completed the intervention without outcome data. There were no statistically significant baseline differences between patients who completed the study and those who dropped out. Of the 108 patients who completed the study, 40 (37%) patients were men, and 68 (63%) patients were women, with a mean age of 58.6 (SD 9.3; range 34-74) years (Table 1). Most patients lived with someone (93/108, 86.1%), were highly educated (87/108, 80.6%), and were in a partnership (78/108, 72.2%). Patients were instructed to use the NEVERMIND system for a total of 12 weeks. However, the average number of days of use in the NEVERMIND study was 44.9 days, which is approximately 6 weeks, and only 12 patients had used the NEVERMIND system for the recommended period of ≥12 weeks (data not shown).

Table 1. Multivariate regression of sociodemographic characteristics and the System Usability Scale (SUS) score at 4 weeks (interim) and 12 weeks (final).

	Participants, n (%)	SUS (interim)		SUS (final)	
		Value, mean (SD)[a]	β coefficient (95% CI)	Value, mean (SD)[b]	β coefficient (95% CI)
Total	108 (100)	70.9 (12.3)	N/A[c]	73.4 (12.5)	N/A
Sex					
Men (reference)	40 (37)	72.01 (11)	−2.88 (−8.6 to 2.8)	72.1 (13.6)	−0.31 (−5.8 to 5.2)
Women	68 (63)	70.4 (13)	N/A	74.2 (11.9)	N/A
Age (years), mean (SD)	58.6 (9.3)	N/A	−0.08 (−0.4 to 0.2)	N/A	−0.24 (−0.5 to 0.1)
Living arrangement					
Living alone (reference)	15 (13.8)	71.1 (14.2)	2.77 (−6.9 to 12.5)	69.5 (11.8)	6.57 (−3.1 to 16.2)
Cohabitant	93 (86.1)	70.9 (12.1)	N/A	74.0 (12.6)	N/A
Education					
Low (reference)	21 (19.4)	67.8 (11.8)	4.64 (−1.7 to 11.0)	70.4 (11.2)	4.26 (−1.96 to 10.5)
High	87 (80.5)	71.7 (12.4)	N/A	74.2 (12.8)	N/A
Marital status					
Single (reference)	30 (25.9)	72.1 (11.8)	−3.66 (−10 to 3.7)	73.2 (11.1)	−3.09 (−10.4 to 4.2)
Married	78 (72.2)	70.5 (12.5)	N/A	73.5 (13.1)	N/A
Employment					
Unemployed (reference)	56 (51.9)	70.2 (11.9)	.30 (−5.3 to 5.9)	72.2 (13.3)	−0.50 (−6.0 to 4.9)
Employed	52 (44.4)	71.7 (12.7)	N/A	74.7 (11.7)	N/A

[a]n=104.

[b]n=107.

[c]N/A: not applicable.

SUS Score

Table 1 shows how different sociodemographic characteristics are related to the SUS score described in a multivariate regression model using the β coefficient and 95% CI. All patients except one (107/108, 99.1%) had data for the SUS at the final follow-up, and 3.7% (4/108) of the patients had missing SUS scores at the interim time point. The mean SUS score at the final time point (mean 73.4, SD 12.5) was higher than the mean SUS score at the interim time point (mean 70.9, SD 12.3; Table 1). However, there were 3 more patients when computing the SUS at the final time point than at the interim time point; the mean SUS score at the final time point, excluding the scores of the patients those who had missing SUS scores at the interim time point, was 73.3 (data not shown).

The mean SUS score at the final time point was normally distributed based on the Skewness and Kurtosis tests (*P*=.05), whereas the mean SUS score at the interim time point was not normally distributed (*P*=.001); thus, a nonparametric regression was appropriate. However, both parametric and nonparametric regressions yielded similar β estimates, with slightly different CIs. Age was significantly associated with SUS score at the final time point in a univariate model (*P*=.04, data not shown), but it became insignificant in a multivariate model after adjusting for other covariates (*P*=.15, data not shown). No other sociodemographic characteristics were associated with a higher SUS score at either the interim or final time points.

uMARS Score

At 12 weeks, 107 patients completed the uMARS. The global uMARS score was 3.8 (Table 2), which is above the average (3.0) for this scale. Each subscale scored above average, with engagement, functionality, aesthetics, and information scoring at 3.5, 3.9, 3.6, and 4.2, respectively (data not shown). Table 2 shows how different sociodemographic characteristics were related to uMARS scores. A Skewness and Kurtosis test showed that the uMARS was normally distributed (*P*=.54, data not shown).

The mean uMARS score was significantly higher for women (mean 3.9, SD 0.3) than that for men (mean 3.7, SD 0.3). No sociodemographic characteristics were associated, either in the univariate analyses or in the multivariate model, with different uMARS scales, except for women rating the app higher than men (*P*=.03; Table 2). A further investigation into the uMARS showed that the subscale *engagement* showed significant differences between women and men (β=0.26; *P*=.02, 95% CI 0.04-0.48). Women had a mean engagement score of 3.64 (SD 0.52; range 2.4-5), whereas men had a mean engagement score of 3.36 (SD 9.51; range 2.2-4.4; data not shown). There were no significant differences in the other subscales.

Table 2. Multivariate regression of sociodemographic characteristics and the user version of the Mobile App Rating Scale score at 12 weeks (final).

Sociodemographic characteristics	β coefficient (SE; 95% CI)	*P* value
Sex	0.16 (0.07; 0.02 to 0.30)	.03[a]
Age (years)	−0.003 (0.004; −0.01 to 0.005)	.43
Marital status	−0.02 (0.09; −0.02 to 0.17)	.83
Living arrangement	0.02 (0.13; −0.23 to 0.27)	.86
Education level	0.02 (0.08; −0.14 to 0.17)	.82
Employment status	−0.07 (0.07; −0.21 to 0.07)	.30

[a]Statistically significant at $P<.05$.

Use of the NEVERMIND System

A total of 99.1% (107/108) of patients had log data for computing the usage index of the NEVERMIND system. The mean usage index was 0.48. The usage index was not normally distributed according to the Skewness and Kurtosis tests ($P<.001$); however, most patients (73/107, 68.2%) had a usage index higher than the mean. However, the distribution was because of a few outliers, and both parametric and nonparametric regressions yielded similar effect estimates. Table 3 shows the results of parametric regression.

No sociodemographic characteristics had statistically significant associations with higher or lower use of the system, except for women, showing a lower usage index than men. The mean usage index for women was 0.43 (SD 0.28; range 0.02-0.99), whereas that for men was 0.56 (SD 0.24; range 0.04-0.97; data not shown). Women had significantly lower use of the NEVERMIND system during the study period.

Table 3. Multivariate regression of sociodemographic characteristics and use of the NEVERMIND (Neurobehavioural Predictive and Personalised Modelling of Depressive Symptoms During Primary Somatic Diseases) system at 12 weeks (final).

Sociodemographic characteristics	β coefficient (SE; 95% CI)	*P* value
Sex	−0.13 (0.06; −0.25 to −0.01)	.04[a]
Age (years)	−0.001 (0.003; −0.01 to 0.01)	.82
Marital status	0.14 (0.08; −0.02 to 0.30)	.09
Living arrangement	−0.11 (0.10; −0.32 to 0.09)	.28
Education level	0.05 (0.07; −0.09 to 0.20)	.49
Employment status	−0.05 (0.06; −0.17 to 0.07)	.42

[a]Statistically significant at $P<.05$.

Discussion

Principal Findings

This study aimed to investigate the different sociodemographic characteristics that can determine higher usability, as measured by 3 usability outcomes. We found that none of the sociodemographic characteristics investigated were associated with different types of usability outcomes except for women rating the mobile app higher on the uMARS and that men having used the system more than women. Several methods exist to measure the usability of systems [12,21]. In our study, we used subjective measurements in tandem with use metrics; this is considered to be a more reliable predictor of use frequency than using subjective scales or logging tasks alone [21].

The patients had a higher SUS score at the final than at the interim time point. The higher SUS score at the final time point could be an experience effect, that is, with more time and opportunity to navigate through the app, patients could have gained app-relevant experience and skills, thereby increasing the app usability. This especially aligns with the SUS, as it comprised 10 statements covering the need for support and training, and complexity of the system—aspects that improve over time.

The SUS enables us to compare our system with other comparable and highly thought of products that serve similar purposes and possibly cater to the same group of users, such as patients affected by other somatic illnesses. For example, Grossert et al [22] reported the usability of a web-based Stress Management Intervention (STREAM) in 11 patients diagnosed with cancer, of whom 4 (36%) were diagnosed with breast cancer and 1 (9%) with prostate cancer. They found the overall SUS score of STREAM to be 83.6, which was higher than the predefined cutoff for good usability and the NEVERMIND system. However, the NEVERMIND system was tested with more patients and was geared mainly toward reducing depressive symptoms. There is a paucity of research on the evaluation of eHealth systems using SUS in patients with cancer in the field of mental health.

uMARS was only administered at the final evaluation, and the global score was recorded as 3.8, with the subscale information scoring very high at 4.2. Being a woman was the only sociodemographic characteristic associated with a higher

uMARS score. Research shows that women have higher health care–seeking behavior, especially when it comes to mental health care [23], which can lead to higher engagement with the mobile app.

The uMARS is also widely used to evaluate different mobile apps geared toward patients diagnosed with cancer. A recent systematic descriptive search conducted by Amor-García et al [24] analyzed 46 apps available for patients diagnosed with different types of cancer, including prostate cancer. They found that the mean Mobile Appl Rating Scale score of these 46 apps was 2.98, with 13 apps scoring ≥3.5. In another evaluation of mobile apps designed for patients diagnosed with cancer, including patients with breast or prostate cancer, Böhme et al [25] reported a significantly lower score (1.96). Interestingly, these previous studies also noted the engagement of patients to be the lowest scored subscale, similar to what we observed in our sample group. As one of the goals of self-management eHealth tools is to increase the engagement of patients in managing their health, more work is needed in this area.

The usage index metric was used as a quantitative usability measure by looking at how many patients used the sensorized shirt and mobile app until the end of the study. A usage index of 1 indicates that the patient has used either the mobile app or the sensorized shirt or both at least once a day for the duration of the study. Similar to the uMARS scale, this study showed that there was a difference of usage between women and men. Our results showed that women used the system less frequently during the study period. Although the system comprised a sensorized shirt and mobile app, patients were instructed to use the shirt twice a week, with the mobile app being intended for everyday use. Research has indicated that women interact with and use specific types of mobile health apps that are geared toward nutrition and self-care, whereas men interact more with physical activity–related mobile health apps [26]. Therefore, the higher use among men might be related to the contents of the NEVERMIND system, which might have a more engaging physical activity module than the dietary recommendation and mindfulness modules.

Limitations

This study had some limitations. First, the sample size used for the study was based on the initial sample size calculated to test the clinical effectiveness of the NEVERMIND system, which made it impossible to include more patients. Consequently, the generalizability of the results of this study is limited with respect to giving a definitive conclusion regarding the association between sociodemographic characteristics and usability. However, our results can provide an indication of how sociodemographic characteristics might be associated with usability, which has been documented in previous research. Furthermore, other variables, such as digital literacy and the ability to use these types of technologies, were not included, which can also influence the generalizability of the results to other populations who might have different starting digital literacy despite having similar sociodemographic characteristics. Our study found that sex was associated with differential uMARS scores and use of the system, which inadvertently fully aligns with the cancer diagnosis; that is, all patients diagnosed with breast cancer were women and all patients diagnosed with prostate cancer were men. However, there was no other information available to differentiate sex from cancer diagnosis. The length of use of the NEVERMIND system is another limitation. Ideally, all patients should have used the system for 12 weeks, which was the recommended time in the NEVERMIND study and might have influenced how patients rated the system. Owing to the small sample size, it was not possible to analyze only those who used the system for 12 weeks or more. Another limitation is the choice of outcome measures. For example, the usage index metric does not provide information about the number of tasks completed, time on a task, or error on a task, all of which are important predictors for usability evaluation in eHealth [27]. Thus, the use metric is only a partial indicator of usability. However, using subjective usability measures, such as the SUS and uMARS, coupled with use metrics, contributes to a better benchmark for usability evaluation.

Conclusions

Research demonstrates that different sociodemographic characteristics are associated with higher use and efficacy of eHealth interventions. Despite the limitations of the study, our initial findings suggest that the usability of the NEVERMIND system does not suffer from a large *digital divide* where certain sociodemographic characteristics are more associated with higher usability. There seems to be an indication that there is higher favorability of the mobile app among women but that men use the NEVERMIND system more. Future research will focus on examining specific modules separately in the NEVERMIND system to understand content-related differences in usability.

Acknowledgments

The NEVERMIND (Neurobehavioural Predictive and Personalised Modelling of Depressive Symptoms During Primary Somatic Diseases) project was funded by the European Union's Horizon 2020 Research and Innovation Programme (grant 689691). The authors would like to thank all researchers and affiliated staff who participated in the implementation of the NEVERMIND project. In addition, the authors would like to thank all the researchers and affiliated staff in the Department of Clinical and Biological Sciences, University of Turin, for recruiting all patients included in this study. All the authors approved the final manuscript.

Authors' Contributions

NGP wrote the first draft, conducted the statistical analysis, and revised the manuscript on the basis of input from other coauthors. GH provided input for the statistical analysis and critically revised several drafts of the manuscript. SC and LO provided data for patients and critically revised all parts of the manuscript. SGM and MO participated in data management and revised some parts of the manuscript. BM critically revised the manuscript and contributed to the statistical analysis. EPS was the principal investigator of the NEVERMIND project and reviewed the final draft of the manuscript. VC provided input on the design of the study and statistical analysis and critically revised several drafts of the manuscript.

Conflicts of Interest

None declared.

References

1. Carreira H, Williams R, Müller M, Harewood R, Stanway S, Bhaskaran K. Associations between breast cancer survivorship and adverse mental health outcomes: a systematic review. J Natl Cancer Inst 2018 Dec 01;110(12):1311-1327 [FREE Full text] [doi: 10.1093/jnci/djy177] [Medline: 30403799]
2. Krebber A, Buffart L, Kleijn G, Riepma I, de Bree R, Leemans C, et al. Prevalence of depression in cancer patients: a meta-analysis of diagnostic interviews and self-report instruments. Psychooncology 2014 Feb;23(2):121-130 [FREE Full text] [doi: 10.1002/pon.3409] [Medline: 24105788]
3. Zhu J, Fang F, Sjölander A, Fall K, Adami H, Valdimarsdóttir U. First-onset mental disorders after cancer diagnosis and cancer-specific mortality: a nationwide cohort study. Ann Oncol 2017 Aug 01;28(8):1964-1969 [FREE Full text] [doi: 10.1093/annonc/mdx265] [Medline: 28525559]
4. Du L, Shi HY, Yu HR, Liu XM, Jin XH, Yan-Qian, et al. Incidence of suicide death in patients with cancer: a systematic review and meta-analysis. J Affect Disord 2020 Nov 01;276:711-719. [doi: 10.1016/j.jad.2020.07.082] [Medline: 32794450]
5. Deady M, Choi I, Calvo R, Glozier N, Christensen H, Harvey S. eHealth interventions for the prevention of depression and anxiety in the general population: a systematic review and meta-analysis. BMC Psychiatry 2017 Aug 29;17(1):310 [FREE Full text] [doi: 10.1186/s12888-017-1473-1] [Medline: 28851342]
6. Apolinário-Hagen J, Vehreschild V, Alkoudmani RM. Current views and perspectives on e-mental health: an exploratory survey study for understanding public attitudes toward internet-based psychotherapy in Germany. JMIR Ment Health 2017 Feb 23;4(1):e8 [FREE Full text] [doi: 10.2196/mental.6375] [Medline: 28232298]
7. How can eHealth improve care for people with multimorbidity in Europe? ICARE4EU. URL: https://www.euro.who.int/__data/assets/pdf_file/0007/337588/PB_25.pdf [accessed 2022-03-09]
8. Naslund J, Marsch L, McHugo G, Bartels S. Emerging mHealth and eHealth interventions for serious mental illness: a review of the literature. J Ment Health 2015;24(5):321-332 [FREE Full text] [doi: 10.3109/09638237.2015.1019054] [Medline: 26017625]
9. Bäuerle A, Teufel M, Schug C, Skoda E, Beckmann M, Schäffeler N, et al. Web-based MINDfulness and Skills-based distress reduction in cancer (MINDS): study protocol for a multicentre observational healthcare study. BMJ Open 2020 Aug 13;10(8):e036466 [FREE Full text] [doi: 10.1136/bmjopen-2019-036466] [Medline: 32792437]
10. Carli V, Wasserman D, Hadlaczky G, Petros N, Carletto S, Citi L, et al. A protocol for a multicentre, parallel-group, pragmatic randomised controlled trial to evaluate the NEVERMIND system in preventing and treating depression in patients with severe somatic conditions. BMC Psychiatry 2020 Mar 02;20(1):93 [FREE Full text] [doi: 10.1186/s12888-020-02494-3] [Medline: 32122315]
11. Enam A, Torres-Bonilla J, Eriksson H. Evidence-based evaluation of eHealth interventions: systematic literature review. J Med Internet Res 2018 Nov 23;20(11):e10971 [FREE Full text] [doi: 10.2196/10971] [Medline: 30470678]
12. Sousa V, Lopez K. Towards usable e-health. Appl Clin Inform 2017 Dec 21;08(02):470-490 [FREE Full text] [doi: 10.4338/aci-2016-10-r-0170]
13. Jokela T, Iivari N, Matero J, Karukka M. The standard of user-centered design and the standard definition of usability: analyzing ISO 13407 against ISO 9241-11. In: Proceedings of the Latin American conference on Human-computer interaction. 2003 Presented at: CLIHC03: Latin American Conference on Human-Computer Interaction; Aug 17 - 20, 2003; Rio de Janeiro Brazil URL: https://dl.acm.org/doi/10.1145/944519.944525 [doi: 10.1145/944519.944525]
14. Reiners F, Sturm J, Bouw L, Wouters E. Sociodemographic factors influencing the use of eHealth in people with chronic diseases. Int J Environ Res Public Health 2019 Feb 21;16(4):645 [FREE Full text] [doi: 10.3390/ijerph16040645] [Medline: 30795623]
15. Carli V, Petros NG, Hadlaczky G, Vitcheva T, Berchialla P, Bianchi S, et al. The NEVERMIND e-health system in the treatment of depressive symptoms among patients with severe somatic conditions: a multicentre, pragmatic randomised controlled trial. EClinicalMedicine 2022 Jun 1;48:101423. [doi: 10.1016/j.eclinm.2022.101423]
16. Brooke J. SUS: A 'Quick and Dirty' Usability Scale. Boca Raton, Florida, United States: CRC Press; 1996.
17. Sauro J. A Practical Guide to the System Usability Scale: Background, Benchmarks & Best Practices. Scotts Valley, California, US : Createspace Independent Pub; 2011.

18. Teo CH, Ng CJ, Lo SK, Lim CD, White A. A mobile web app to improve health screening uptake in men (ScreenMen): utility and usability evaluation study. JMIR Mhealth Uhealth 2019 Apr 15;7(4):e10216 [FREE Full text] [doi: 10.2196/10216] [Medline: 30985280]
19. van der Velde M, Valkenet K, Geleijn E, Kruisselbrink M, Marsman M, Janssen LM, et al. Usability and preliminary effectiveness of a preoperative mHealth app for people undergoing major surgery: pilot randomized controlled trial. JMIR Mhealth Uhealth 2021 Jan 07;9(1):e23402 [FREE Full text] [doi: 10.2196/23402] [Medline: 33410758]
20. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the User Version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016 Jun 10;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]
21. Harrati N, Bouchrika I, Tari A, Ladjailia A. Exploring user satisfaction for e-learning systems via usage-based metrics and system usability scale analysis. Comput Human Behav 2016 Aug;61:463-471 [FREE Full text] [doi: 10.1016/j.chb.2016.03.051]
22. Grossert A, Heinz S, Müller L, Gaab J, Urech C, Berger TH, et al. Usability testing of the online stress management intervention (STREAM) for cancer patients: results and implementations. Cancer Stud Ther J 2019 May 13;4(2). [doi: 10.31038/CST.2019422]
23. Thompson A, Anisimowicz Y, Miedema B, Hogg W, Wodchis W, Aubrey-Bassler K. The influence of gender and other patient characteristics on health care-seeking behaviour: a QUALICOPC study. BMC Fam Pract 2016 Mar 31;17:38 [FREE Full text] [doi: 10.1186/s12875-016-0440-0] [Medline: 27036116]
24. Amor-García M, Collado-Borrell R, Escudero-Vilaplana V, Melgarejo-Ortuño A, Herranz-Alonso A, Arija A, et al. Assessing apps for patients with genitourinary tumors using the mobile application rating scale (MARS): systematic search in app stores and content analysis. JMIR Mhealth Uhealth 2020 Jul 23;8(7):e17609 [FREE Full text] [doi: 10.2196/17609] [Medline: 32706737]
25. Böhme C, von Osthoff MB, Frey K, Hübner J. Qualitative evaluation of mobile cancer apps with particular attention to the target group, content, and advertising. J Cancer Res Clin Oncol 2018 Jan;144(1):173-181 [FREE Full text] [doi: 10.1007/s00432-017-2533-0] [Medline: 29052039]
26. Bol N, Helberger N, Weert J. Differences in mobile health app use: a source of new digital inequalities? Inf Soc 2018 Apr 26;34(3):183-193 [FREE Full text] [doi: 10.1080/01972243.2018.1438550]
27. Broekhuis M, van Velsen L, Hermens H. Assessing usability of eHealth technology: a comparison of usability benchmarking instruments. Int J Med Inform 2019 Aug;128:24-31. [doi: 10.1016/j.ijmedinf.2019.05.001] [Medline: 31160008]

Abbreviations

NEVERMIND: Neurobehavioural Predictive and Personalised Modelling of Depressive Symptoms During Primary Somatic Diseases
SUS: System Usability Scale
uMARS: user version of the Mobile App Rating Scale

Edited by A Mavragani; submitted 14.10.21; peer-reviewed by O Blanson Henkemans, L Costantini; comments to author 09.01.22; revised version received 09.03.22; accepted 18.03.22; published 08.06.22.

Please cite as:
Petros NG, Hadlaczky G, Carletto S, Martinez SG, Ostacoli L, Ottaviano M, Meyer B, Scilingo EP, Carli V
Sociodemographic Characteristics Associated With an eHealth System Designed to Reduce Depressive Symptoms Among Patients With Breast or Prostate Cancer: Prospective Study
JMIR Form Res 2022;6(6):e33734
URL: https://formative.jmir.org/2022/6/e33734
doi:10.2196/33734
PMID:35675116

Original Paper

Feasibility, Acceptability, and Preliminary Efficacy of an App-Based Meditation Intervention to Decrease Firefighter Psychological Distress and Burnout: A One-Group Pilot Study

Thaddeus W W Pace[1,2,3], PhD; Katharine H Zeiders[4], PhD; Stephanie H Cook[5,6], PhD; Evelyn D Sarsar[4], MA; Lindsay T Hoyt[7], PhD; Nicholas L Mirin[5], MPH; Erica P Wood[5], MPH; Raquel Tatar[8], PhD; Richard J Davidson[9], PhD

[1]Division of Biobehavioral Health Science, College of Nursing, University of Arizona, Tucson, AZ, United States
[2]Department of Psychiatry, College of Medicine, University of Arizona, Tucson, AZ, United States
[3]Department of Psychology, College of Science, University of Arizona, Tucson, AZ, United States
[4]Norton School of Family and Consumer Sciences, College of Agriculture and Life Sciences, University of Arizona, Tucson, AZ, United States
[5]Department of Social and Behavioral Sciences, School of Global Public Health, New York University, New York, NY, United States
[6]Department of Biostatistics, School of Global Public Health, New York University, New York, NY, United States
[7]Department of Psychology, Fordham University, Bronx, NY, United States
[8]Healthy Minds Innovation, Madison, WI, United States
[9]Department of Psychology, University of Wisconsin, Madison, WI, United States

Corresponding Author:
Thaddeus W W Pace, PhD
Division of Biobehavioral Health Science
College of Nursing
University of Arizona
1305 N Martin Ave
Tucson, AZ, 85721
United States
Phone: 1 520 626 3520
Email: twwpace@arizona.edu

Abstract

Background: Firefighters are often exposed to occupational stressors that can result in psychological distress (ie, anxiety and depression) and burnout. These occupational stressors have only intensified with the onset of the COVID-19 pandemic and will likely persist in the postpandemic world.

Objective: To address occupational stressors confronting firefighters, we pilot tested a novel, cost-effective, smartphone app–based meditation intervention created by Healthy Minds Innovations that focused on mindfulness (awareness) training along with practices designed to cultivate positive relationships (connection), insight into the nature of the self (insight), and a sense of purpose in the context of challenge (purpose) with a sample of professional firefighters from a large metropolitan area in southwestern United States.

Methods: A total of 35 participants were recruited from a closed online group listserv and completed the self-guided 10-unit meditation app over the course of 10 days, at 1 unit per day. We assessed anxiety symptoms, depression symptoms, burnout, and negative affect as well as saliva diurnal cortisol rhythm, an objective indicator of stress-related biology, before and after use of the meditation app.

Results: This study demonstrated the meditation app was both feasible and acceptable for use by the majority of firefighters. We also found significant reductions in firefighters' anxiety (P=.01), burnout (P=.05), and negative affect (P=.04), as well as changes in cortisol diurnal rhythm, such as waking cortisol (P=.02), from before to after use of the meditation app.

Conclusions: Our study findings call for future research to demonstrate the efficacy of this meditation app to reduce psychological distress and burnout in firefighters.

(JMIR Form Res 2022;6(6):e34951) doi:10.2196/34951

KEYWORDS

firefighter; meditation; smartphone app; anxiety; cortisol; digital health; mobile health; mHealth; mental health; burnout; stress management

Introduction

Along with other first responders, firefighters are often exposed to occupational stressors such as interpersonal conflict and workplace fairness issues [1], as well as resuscitations and other clinical emergencies that can be psychologically traumatic. Exposure to these occupational stressors increases the risk that firefighters will develop stress-related chronic illnesses, such as anxiety and depression, posttraumatic stress disorder (PTSD), and chronic pain [2]. Throughout the COVID-19 pandemic, occupational stressors that drive the risk for stress-related illnesses in firefighters have only intensified [3,4] and will likely continue in the postpandemic world. To address psychological distress (ie, anxiety and depression) resulting from exposure to occupational stressors confronting firefighters, we pilot tested a novel, cost-effective, smartphone app–based meditation intervention that focused on mindfulness, connection with others, and compassion for the self and others.

The term "meditation" refers to contemplative practices that bring mental capacities and processes under greater voluntary control not just during but also between practice sessions [5,6]. Mindfulness is a popular style of meditation that is designed to cultivate the quality of nonjudgmental awareness in the present moment [6,7]. The benefits of structured mindfulness meditation interventions to reduce anxiety and depression, as well as to improve various stress-related biomarkers, have been demonstrated in several different populations, ranging from adolescents to cancer survivors to police officers [8-10]. A few published studies suggest that mindfulness meditation may also improve various aspects of health for firefighters [11-14], and 2 studies, both with the same intervention, tested an intervention delivered to firefighters virtually (ie, not in person) [13,15]. There has also been descriptive research on dispositional mindfulness in firefighters that supports the use of mindfulness meditation in this group. Specifically, dispositional mindfulness (or awareness) has been inversely associated with firefighters' PTSD symptoms, features of anxiety and depression, perceived stress, and suicide risk [7,14,16,17]. Together, these findings suggest that interventions intended to cultivate mindfulness and awareness in firefighters, including by way of meditation, may be effective to reduce their anxiety and depression.

Besides mindfulness, available evidence suggests that other styles of meditation may be worthwhile for firefighters. One group of structured meditation interventions collectively called "compassion meditation" (eg, Cognitively-Based Compassion Training) are designed to promote compassion for the self and others. Previous research has demonstrated that compassion meditation interventions improve anxiety, depression, and perceived stress in different populations [18-20]. Although we are not aware of any studies to date that have explored the benefits of compassion meditation for firefighters, perceived social connection (ie, social support) has been found to buffer against the effects of trauma exposure and PTSD symptoms in firefighters [21]. This suggests that firefighters may benefit from a meditation intervention that actively cultivates the perception of the importance of connection to others in addition to nonjudgmental awareness in the present moment (ie, mindfulness). We therefore tested a meditation app intervention that included both mindfulness and compassion. Our selection of content for the app was guided by Lazarus and Folkman's theory of stress and coping [22], which has been used previously to study health in firefighters [23], as well as research indicating the importance of mindfulness and perceived social connection for firefighter health [7,14,16,17,21].

We decided to study an asynchronous app-based approach instead of an in-person meditation intervention because conditions during the COVID-19 pandemic (when the study was conducted) prevented large gatherings of people. We also selected an app-based approach to reduce intervention costs. While in-person meditation interventions with trained interventionists can be relatively costly (ie, approximately US $60 per person per group session), asynchronous app-based interventions have fixed development costs that can be recouped over time.

Guided by the stress and coping theory as well as prior research on how stress affects the health of firefighters [1,7,14,16,17], we assessed self-reported anxiety symptoms, depression symptoms, burnout, and negative affect before and after the 10-day meditation intervention. We also included an objective biological outcome of stress, diurnal cortisol rhythm, which is the profile of the cortisol concentration in saliva over the course of the day that is normally high in the morning, peaks 30-60 minutes after wakening (cortisol awakening response [CAR]), and is low at night, indicating healthy hypothalamic-pituitary adrenal (HPA) axis function [24]. Relative disruptions in diurnal cortisol rhythm (eg, high bedtime levels, a flatter diurnal slope from waking to bedtime, high overall concentration across the day) have been associated with self-reported stress and other negative psychological health outcomes (eg, depression) [24,25]. Meditation interventions have been found in prior research to improve or stabilize diurnal cortisol rhythm [26,27], especially in those at risk after exposure to stress [28]. Besides predicting that the meditation app would be feasible to implement with firefighters, we hypothesized that symptoms of anxiety, depression, burnout, and negative affect would decline from before to after use of the meditation app. We also hypothesized that changes in diurnal cortisol rhythm, suggestive of improved HPA axis function, would be evident after use of the meditation app.

Methods

Participants

Participants were career firefighters from a large metropolitan city in southwestern United States (N=35) who were recruited via flyers emailed to a firefighters' union closed listserv. To be eligible for the study, participants had to (1) be actively working as a firefighter in the geographic region of interest; (2) work a

typical 24 hours on/48 hours off schedule; (3) have a smartphone with requisite knowledge to operate their device, including for internet access; (4) have no vacation time scheduled for the duration of the study; (5) report no current symptoms of COVID-19; and (6) not be taking any corticosteroid medication during the study period. Interested firefighters completed an online survey to verify their eligibility and provided informed consent before enrollment and after being informed about the length and purpose of the study in real time by the study staff.

Ethics Approval

All procedures for this research were approved by the University of Arizona Human Subjects Protection Program (Project #2005659631).

Design and Procedures

This study used a 1-group pretest-posttest design. After giving consent online, participants were contacted by study personnel via telephone to discuss the study protocol and the study start date. Participants were then sent a study packet via mail that included an information sheet describing the daily activities, saliva sampling protocol, instructions on downloading the meditation app, and a "spit kit" to collect saliva samples, which included 6 vials (with labels to record sample collection time), straws, and prepaid postage to send samples back to the laboratory.

The study protocol was conducted over the course of 16 days, including preassessments, intervention, and postassessments. To aid with compliance to the study protocol, study personnel texted or called participants each day of the study to remind them of the daily tasks and answer any concerns or questions. Participants completed aspects of the study, including data collection, from their workplace or at home from on smartphones. On day 1, participants completed a closed online baseline survey via Qualtrics software that assessed demographic characteristics and psychosocial and behavioral constructs. On days 2 and 3, participants completed a brief daily online survey in the mornings that included self-report assessments (eg, anxiety and depression), COVID-19-related stress, and sleep quality (findings regarding the latter two are reported elsewhere). They also provided saliva samples upon waking, 30 minutes after waking, and at bedtime on both days (6 total samples). On days 4 to 13 (10 days), participants completed the 10-minute meditation segments guided by the app and a brief online daily survey (both administered via their smartphones). On days 14 and 15, participants ended use of the meditation app but continued with the online self-report data collection and completed 2 more additional days of saliva sampling (upon waking, 30 minutes after waking, and at bedtime, with a total of 6 samples). On day 16, participants completed a final online survey. Around this time, participants were also contacted by study staff to gather comments about the meditation app and the study procedures. Participants were then instructed to ship their samples to the research team using the overnight return envelopes provided by the study team. Participants were compensated US $200 if they completed over 60% of the daily online surveys and US $160 if they completed less than 60%. Study information was stored on password-protected servers that were only accessible by the research team.

Intervention

The smartphone app–based meditation intervention we tested was created by Healthy Minds Innovations (HMI) [29]. The app consists of 10 individual 10-minutes sessions, with 1 session per day conducted over 10 consecutive days. Our selection of content in the version of the app for this study with firefighters was guided by Lazarus and Folkman's theory of stress and coping [22], which has been used previously to guide a firefighter study [23], as well as research indicating the importance of mindfulness and perceived social connection for firefighter health [7,14,16,17,21]. Accordingly, the app provided audio recordings on the constructs of awareness (or mindfulness) and connection (with others/compassion). The app also included sessions on insight (including compassion for the self and others) and purpose (eg, finding meaning in challenges experienced by the self and others), which collectively promoted lessons of connection. The app remained frozen and unchanged during the study period. Screenshots of the app homes screen and an app session are shown in Figure 1.

After enrolling and completing the initial assessment, study staff provided a link and unique access code to each participant to download the meditation app created by HMI to their phone (either iPhone or Android). After accessing the meditation app, participants were presented with a welcome screen, followed by a screen that presented them with the option to turn on a notification allowing the app to remind them to practice at a particular time of day, if desired. Participants then landed on a "meditations" page, from which they were able to select the day 1 module. Including the first day, the app contained 10 individual modules that covered the overarching constructs of awareness, connection, insight, and purpose. A description of each module by day is presented in Table 1. Participants were able to select a module only after they completed the prior module. Participants were instructed to complete 1 module per day over the course of the 10-day intervention period but were allowed to catch up on units if they missed a day.

Figure 1. The Healthy Minds Innovations app home screen (A) and a session screen (B).

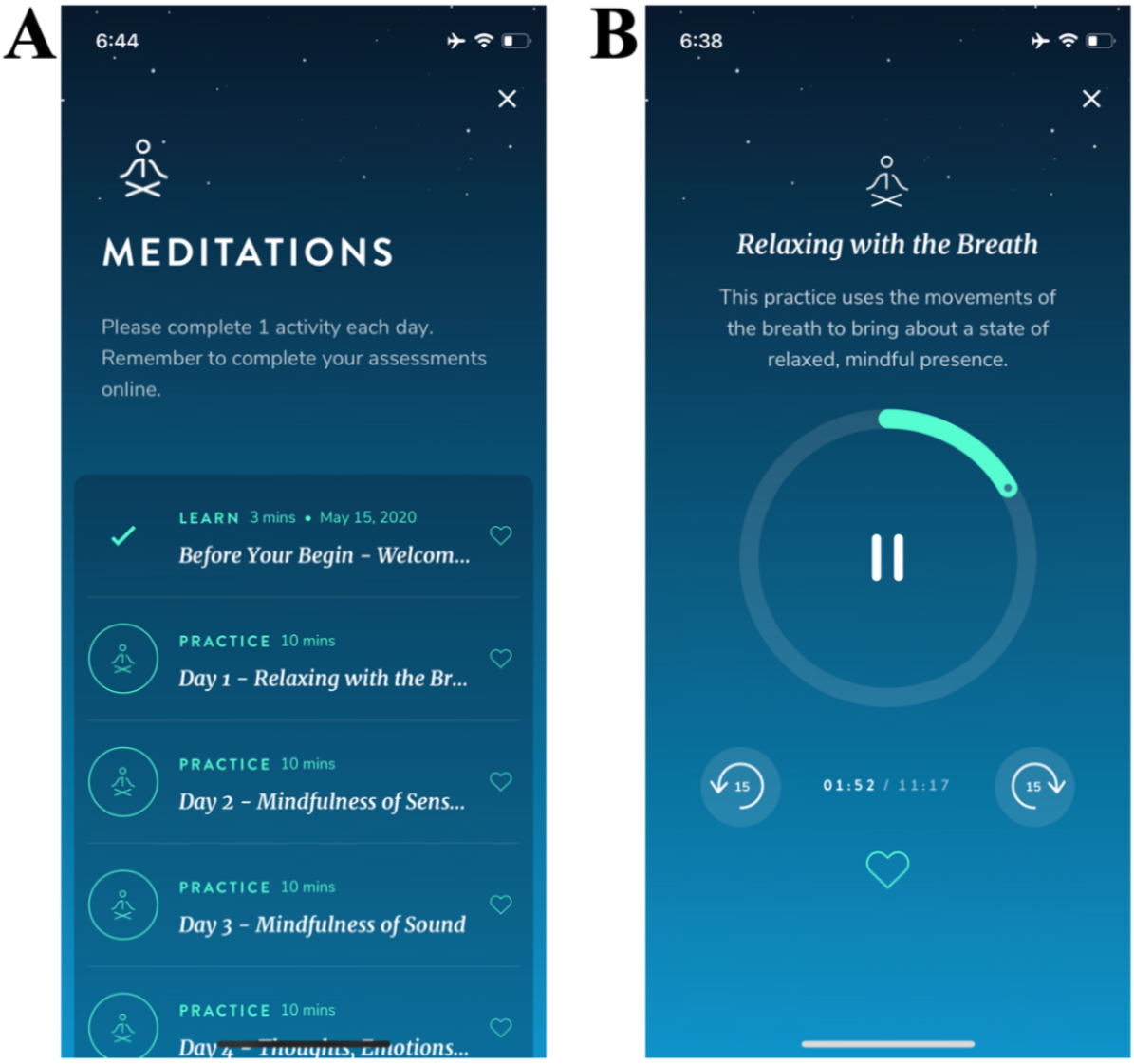


Table 1. The Healthy Minds Innovation app with construct focus and description by day.

Day	Topic title	Construct focus	Description
1	Relaxing with the breath	Awareness	Attending to the breath while staying in the present moment with the body
2	Mindfulness of sensations	Awareness	Attending to sensations while in the present moment with the body
3	Mindfulness of sound	Awareness	Attending to sounds while staying in the present moment
4	Thoughts, emotions, and the breath	Awareness	Attending to thoughts, emotions, and breath while staying in the present moment with the body
5	Seeing the good in ourselves	Connection	Shifting awareness of the body to our natural talents and strengths
6	Gratitude	Connection	Realizing and giving thanks for our connections to others
7	Compassion in difficult situations	Connection	Realizing our innate wish to be happy and free from suffering and extending that wish to others
8	Question your assumptions	Insight	Examining our unconscious beliefs that influence how we see ourselves and how we view other people and situations around us
9	Values in difficult times	Purpose	Applying our values in the moment to help us stay grounded and resilient
10	The meaning of adversity	Purpose	Finding meaning in challenges experienced by us and others

Quantitative Outcomes

Self-report instruments containing 46 individual items were tested extensively by 2 members of the research team (authors NLM and EPW) for usability and technical functionality in Qualtrics before being administered to the participants. Qualtrics evaluated the responses to all items for completeness before participants were able to select the "finished" button at the bottom of the questionnaires screen. Participants were also able to review their responses before clicking "finished." Each enrolled participant was allowed to respond to the set of items only once at each assessment time point, which was managed by Qualtrics using cookies. The university logo was displayed at the top of the Qualtrics screen.

Anxiety Symptoms

The Patient Reported Outcomes Measurement Information System (PROMIS) Short Form v1.0– Emotional Distress–Anxiety 8a was used at baseline and at the end of the intervention to assess firefighters' experiences of emotions such as fear, stress, and anxiety within the past 7 days. This scale was developed with the Item Response Theory [30]. Responses were rated on a 5-point Likert scale ranging from 1 (never) to 5 (always), and a total raw score was obtained by summing the scores of all items. Summative raw scores can range from 8 to 40 and are converted to *t* scores to compare to population norms, with higher scores indicating greater anxiety symptoms. This scale demonstrated adequate reliability at the baseline (α=.91) and end (α=.88) assessments.

Depressive Symptoms

PROMIS Short Form v1.0–Emotional Distress–Depression–Short Form 8a was used at baseline and at the end of the intervention to assess the frequency of firefighters' experiences of emotions such as worthlessness, hopelessness, and sadness. This scale was also developed with the Item Response Theory [30], and responses were rated on a 5-point Likert scale ranging from 1 (never) to 5 (always). The total raw score is obtained by summing the scores of all items ranging from 8 to 40 and converting them to *t* scores to compare to population norms, with higher scores indicating greater depressive symptoms. This scale demonstrated adequate reliability at the baseline (α=.92) and end (α=.87) assessments.

Burnout

The short version of the 10-item Burnout Measure was used to assess firefighters' symptoms of burnout at baseline and at the end of the intervention [31]. The scale asks participants to rate their feelings about their work (eg, tired, disappointed with people, hopeless), and responses are rated on a 7-point Likert scale ranging from 1 (never) to 7 (always). A score of 0 to 2.4 indicates a very low level of burnout, a score between 2.5 and 3.4 indicates danger signs of burnout, and a score between 3.5 and 4.4 indicates burnout. The scale has been used with firefighters in prior work [32], and in this study, demonstrated adequate reliability at the baseline (α=.86) and end (α=.84) assessments.

Negative Affect

The Positive and Negative Affect Schedule (PANAS-SF) [33] was used assess firefighters' negative affect on at baseline and at the end of the intervention. Participants were asked to respond to 10 items (eg, distressed, upset) using a 5-point Likert scale ranging from 1 (very slightly or not at all) to 5 (extremely). Scores range from 10 to 50, with higher scores reflecting higher levels of affect in either domain. For negative affect, the internal

consistency was shown to be adequately reliable for day 2 (α=.84) and day 14 (α=.82).

Salivary Cortisol

Saliva was gathered for 2 consecutive days before the intervention (days 2 and 3) and after the intervention (days 14 and 15); on each sampling day, participants provided samples upon waking, 30 minutes after waking, and at bedtime. Participants passively drooled through a straw into a 2 mL polypropylene tube and labeled each tube with the time and date. Participants were instructed not to eat, drink, or brush their teeth 30 minutes before each sample and to keep completed samples in the refrigerator throughout the study period. Participants shipped the saliva samples to the lab, which were then stored at −20 C. Saliva samples were batch assayed in duplicate according to manufacturer instructions for cortisol concentrations using an enzyme immunoassay kit (Salimetrics). Inter- and intra-assay coefficients of variability were 8.16% and 7.74%, respectively. In line with prior work [34], the following cortisol parameters were assessed: waking level (sample 1), bedtime level (sample 3), CAR (the difference between sample 2 and sample 1), diurnal slope (the difference between sample 3 and sample 1), and area under the curve (AUC) to assess total cortisol output across the day. AUC was calculated using the trapezoidal method.

Given that diurnal cortisol patterns are impacted by sleep quality and night awakening [35], firefighters were scheduled to begin the study so that their first day of cortisol sampling occurred on their second day off from work, and their second day of sampling occurred on their first day on work. This eliminated firefighters having to take a waking sample following a shift (for which a waking sample would be hard to determine given that most firefighters are woken up multiple times in the night due to emergency calls).

Determination of Feasibility and Statistical Analytic Plan

Intervention feasibility was defined as 80% or more of the participants completing 9 or more of the 10 meditation app sessions over the 10-day intervention period. Acceptability was determined by comments about the app and study procedures derived from poststudy interviews. Feasibility to collect saliva samples and self-report data was defined as collection of 80% or more of saliva samples or self-report variables at appropriate time points, according to the study protocol. Pre- to postintervention changes in study outcomes were analyzed with pairwise comparison *t* tests for self-reported (ie, burnout, anxiety symptoms, depressive symptoms, negative affect) and cortisol measures.

Results

Participants

The entire study protocol was conducted between June and July 2020. Majority of the firefighters identified as men (n=30, 86%) and were 37 years old on average (SD 8.04). Firefighters identified their racial/ethnic background as White (n=22, 63%); Hispanic or Latino (n=8, 23%); Black or African American (n=1, 3%); Native American, American Indian, or Alaskan native (n=1, 3%); and multiracial/ethnic (n=1, 3%). Two of the 35 firefighters identified their racial/ethnic background as other (ie, human, American). On average, the participants had been working as firefighters for 9.8 years (SD 7.58). In terms of education, the participants reported some college, vocational, or technical school (n=6, 17%); associate's degree (n=11, 31%), bachelor's degree (n=13, 37%), and some advanced work or a master's degree (n=5, 14%). Most firefighters reported their relationship status as married (n=23, 66%), while 9% (n=3) were single, 9% (n=3) divorced, and 17% (n=6) in a relationship but not married (Table 2).

Table 2. Participants' demographic information (N=35).

Participant characteristics	Values
Gender,n (%)	
Men	30 (86)
Women	5 (14)
Ethnic/racial identification, n (%)	
White	22 (63)
Hispanic or Latino	8 (23)
Black or African American	1 (3)
Native American, American Indian, or Alaskan Native	1 (3)
Multiracial/ethnic	1 (3)
Other	2 (6)
Education, n (%)	
Some college	(48)
Bachelor's degree	13 (37)
Advanced (eg, Master's, PhD)	5 (14)
Marital status, n (%)	
Married	23 (66)
Single	3 (9)
Divorced	3 (9)
Age (years), mean (SD)	36.7 (8.03)
Firefighting experience (years), mean (SD)	9.81 (7.57)

Feasibility and Acceptability

Study compliance was high; 33 (94%) of the 35 participants completed the postintervention self-report assessments. As for intervention engagement, 80% (n=28) of the participants completed either 9 or all 10 of the app segments; the remaining participants completed 8 segments (n=2), 5 segments (n=1), 2 segments (n=2), and 1 segment (n = 2). Additionally, 94% (n=33) of the participants collected at least one home saliva sample at both the initial assessment and the end assessment. In poststudy interviews, multiple participants reported that they found use of the app worthwhile. No participants reported concerns with the app or other study procedures. No technical problems or privacy breaches occurred during the study.

Given the interest in preliminary intervention effects, we performed attrition analyses to determine if participants who completed 9 or 10 segments (n=28) differed from those who completed less than 9 (n=8) on relevant demographic characteristics and baseline indicators. No differences emerged for race/ethnicity, gender, relationships status, education level, or age; however, firefighters who completed 9 or 10 HMI segments had been working for more years (mean 10.92, SD 8) than those who completed fewer than 9 HMI segments (mean 5.36, SD 2.87; $t_{28.28}$=2.99, P=.006). No differences emerged on baseline measures of burnout, anxiety and depressive symptoms, negative affect, bedtime cortisol, cortisol awakening response, diurnal slope, or cortisol AUC. There were differences on waking cortisol (t_{31}=−2.67, P=.01); those who participated in most HMI segments had lower waking cortisol levels (mean 0.25, SD 0.10) than those who completed fewer HMI segments (mean 0.38, SD 0.14) at the preintervention assessment.

For the baseline and end in-home saliva sampling protocols, most participants aligned with the schedule ($n_{baseline}$=22; n_{end}=19). There were, however, firefighters whose schedules did not align with that of the study protocol. All firefighters' schedules were on a 24-hour work/48-hour off rotation; however, it is common for firefighters to pick up additional overtime shifts, so they may only have 1 day off in between. If firefighters worked both cortisol sampling days ($n_{baseline}$=2) or if they misreported their work schedules ($n_{baseline}$=2), their cortisol data were not used. If firefighters worked the day prior to the first day of saliva sampling and worked during the second day of saliva sampling, only their second day of samples were used ($n_{baseline}$=5, n_{end}=2). Similarly, if firefighters worked the day prior to their first day of sampling but were off on the second day of sampling, only their second day of samples were used (n_{end}=2). If firefighters were off the day prior to the first day of saliva sampling but worked on both saliva sampling days, only the first day's samples were used ($n_{baseline}$=1, n_{end}=1). Table 3 shows the final number of participants for each of the cortisol outcomes.

Table 3. Pairwise comparison *t* tests from pre- to postintervention.

Outcomes	Sample size, n	Preintervention, mean (SD)	Postintervention, mean (SD)	Pairwise *t* test (*P* value)	Effect size (*d*)
Self-reported outcomes					
Burnout	33	2.48 (0.79)	2.29 (0.69)	t_{32}=2.03 (*P*=.05)	0.35
Anxiety symptoms	33	53.00 (7.99)	49.65 (7.27)	t_{32}=2.70 (*P*=.01)	0.47
Depressive symptoms	33	48.40 (7.65)	47.23 (7.23)	t_{32}=1.35 (*P*=.19)	0.24
Negative affect	26	13.88 (4.48)	12.65 (3.74)	t_{25}=2.17 (*P*=.04)	0.43
Physiological outcomes					
Waking cortisol (µg/dL)	19	0.288 (0.120)	0.221 (0.084)	t_{18}=2.61 (*P*=.02)	0.60
Bedtime cortisol (µg/dL)	19	0.049 (0.053)	0.042 (0.060)	t_{18}=0.34 (*P*=.74)	0.08
CAR[a]	17	0.091 (0.143)	0.082 (0.101)	t_{16}=0.32 (*P*=.75)	0.08
Cortisol diurnal slope	16	−0.016 (0.010)	−0.012 (0.007)	t_{15}=−.1.31 (*P*=.21)	0.33
Cortisol AUC[b]	16	2.60 (1.21)	2.11 (0.63)	t_{15}=2.127 (*P*=.05)	0.53

[a]CAR: cortisol awakening response.

[b]AUC: area under the curve.

Quantitative Outcomes

To examine pre- to postintervention change in outcomes, pairwise comparison *t* tests were performed on self-report (ie, burnout, anxiety symptoms, depressive symptoms, negative affect) and physiological measures (ie, waking cortisol, bedtime cortisol, CAR, diurnal slope, cortisol AUC; Table 3). Significant pre- to postintervention differences were found for burnout, anxiety symptoms, and negative affect (Table 3, Multimedia Appendix 1). In particular, firefighters reported lower burnout, lower anxiety symptoms, and lower negative affect postintervention. Pre- to postintervention differences were not found for depressive symptoms. Significant pre- and posttest differences were also found for measures of waking cortisol and cortisol AUC (Table 3, Multimedia Appendix 1), with firefighters exhibiting lower waking cortisol and lower cortisol AUC after the intervention. There were no pre- and postintervention differences in bedtime cortisol, CAR, or diurnal slope.

Discussion

Principal Results

The goal of this study was to determine the feasibility and acceptability of a novel smartphone-based meditation app for use by firefighters. In addition, we hypothesized that symptoms of anxiety, depression, burnout, and negative affect would decline after use of the app. We also hypothesized that changes in diurnal cortisol rhythm, suggestive of improved HPA axis function, would result from before to after use of the meditation app. We found that use of the app was both feasible and acceptable by a sample of career firefighters. We were also able to collect reliable psychological self-report outcomes and saliva samples from most of our sample, despite the highly variable shift work schedule. In terms of pre- and postintervention outcomes, we found a reduction in self-reported anxiety and burnout symptoms, as well as negative affect. We also observed statistically meaningful (or strong trend) reductions in several measures related to diurnal cortisol rhythm, including waking saliva concentrations of cortisol and cortisol area under the curve, over the course of the day. Together, these findings suggest that use of the HMI smartphone-based meditation app may be associated with improvement in objective stress-related physiological indicators and self-reported psychological factors. We believe these findings, although preliminary, are noteworthy because they provide evidence of the utility and accessibility of a cost-effective app-based meditation intervention to reduce distress and burnout in firefighters.

Limitations

This study has several limitations. First, the sample was relatively small, although it was appropriately sized to investigate feasibility and acceptability, and the power was adequate for the outcomes assessed [36]. Second, the study only included a single arm, with no waitlist control or other comparison group, such as an active attention control. Future research with the meditation app for firefighters should include an appropriate active comparison group that controls for likely nonspecific intervention elements (eg, learning something new from a smartphone app over a period of 10 days) to further establish the efficacy of the meditation app for firefighters. Future studies may also work to determine whether daily text reminders to engage with the app have an impact on results. Third, participants were from the same large metropolitan area. Future studies should involve firefighters from both smaller and larger urban settings, as well as rural settings, to ensure generalizability of findings regarding meditation app efficacy in different regions. Finally, as this was an initial study, we did not control for multiple comparisons. Future studies with larger sample sizes and preregistered hypotheses can address this limitation.

Comparison With Prior Work

Firefighters typically experience occupational stressors that include not only events related to fighting fires but also interpersonal conflict, workplace fairness issues, resuscitations, and other clinical emergencies [1]. Events involving emergency medical care are likely to be encountered more frequently by firefighters than fighting fires, and in some cases, these may be experienced together (eg, when working automobile accidents). Such events can be psychologically traumatic and, along with interpersonal stressors that generate occupational stress, can activate stress-related biological pathways. Repeated activation over time of these pathways by stressor exposure may increase the risk for stress-related chronic illnesses, including diseases that are leading causes of morbidity and mortality in firefighters (eg, cardiovascular disease) [37]. Indeed, a recent systematic review found that occupational stressors experienced by firefighters are associated with various illness states, including PTSD, hypertension, and musculoskeletal pain, as well as objective markers of stress-related biology (ie, heart rate variability) [1]. In this study, we found that use of the meditation app was associated with a decline in self-reported anxiety, burnout, and negative affect, as well as changes in diurnal cortisol rhythm (ie, lower morning cortisol and cortisol AUC), suggesting that the app may reduce the effects of occupational stressors on stress biological pathways in firefighters.

Anxiety is often comorbid with PTSD [38], and as such, any beneficial effects of the meditation app on anxiety may also have a beneficial effect on PTSD symptom severity experienced by many firefighters. We also observed an effect on firefighters' burnout, suggesting that the meditation app may lessen the effect of occupational stressor exposures on firefighters. When firefighters take time for reflection, self-compassion, and connection, they can be more present and positive about their work. Finally, the decline in negative affect we observed was also notable because it highlights how the meditation app may exert a beneficial effect on daily emotional processes, and by extension, on chronic health conditions that may be exacerbated by negative affect [39].

While prior studies have explored the potential benefits of meditation interventions for firefighters, these studies have focused almost exclusively on mindfulness meditation [11-13,15]. Mindfulness meditation is notably different from other forms of meditation, including the HMI meditation app that we studied here that cultivates both mindfulness and feelings of social connection. Although a meditation intervention created by Joyce and colleagues [13,15] called Resilience@Work Mindfulness includes some components of compassion, these components represent only about a sixth of Resilience@Work Mindfulness and only 3 out of 18 of its guided practice audio tracks. This contrast with the HMI meditation app, which includes 5 out of 10 units that involve compassion for the self and/or others. While the HMI app does incorporate mindfulness and attention training, these skills are developed so they can be leveraged in later units to address analytic concepts of compassion for self and connection with others. Thus, the content of the HMI meditation app is distinctly different from mindfulness meditations studied before with firefighters [11-13,15]. In addition, the HMI meditation app is smartphone based, while previous efforts with firefighters have been limited to in-person meditation interventions [11,12] or tablet or computer-facilitated interventions [13,15].

Along with psychological distress, many firefighters experience physical symptoms related to chronic pain that are secondary to musculoskeletal disorders [2], sometimes due to occupational injuries. Although we did not assess pain or other somatic conditions in this study, future studies with this meditation app would do well to include measures of physical health along with distress. While there are many reasons why firefighters experience occupational injuries, the available evidence suggests that injury risk for firefighters may be predicted by psychological distress (ie, depression) [40-43]. The association between occupational injury and psychological distress in firefighters may involve effects of stress exposure on depression and anxiety [44], which in turn may impact awareness and other aspects of dispositional mindfulness [45-47] and therefore workplace safety [48-50].

Conclusions

In this study, we demonstrated the feasibility and acceptability of a 10-day meditation app created by HMI in a sample of career firefighters. We also found that anxiety, burnout, and negative affect improved from before to after use of the meditation app, and we noted changes in various indicators of cortisol diurnal rhythm. These findings suggest that a meditation app with high potential for widespread distribution may improve firefighter health at a low cost. Additional research is needed to demonstrate the efficacy of this intervention versus an active attention control and to show how the meditation app can positively impact other aspects of health that are relevant for firefighters, including physical health and occupational injury risk.

Acknowledgments

KHZ received funding for this study by an intramural research grant from the University of Arizona College of Agriculture and Life Sciences.

Conflicts of Interest

RJD is the founder and president of Healthy Minds Innovations (HMI), a nonprofit organization with a mission to create tools to measure and cultivate well-being. He receives no financial compensation from HMI. RT is the Sr. Director of Measures and Research at HMI. The other authors have conflicts of interest to disclose.

Multimedia Appendix 1
Burnout (panel A), anxiety (panel B), negative affect (panel C), and 2 indicators of diurnal cortisol rhythm (waking saliva concentrations of cortisol [Panel D] and cortisol area under the curve [Panel E]) decreased from before to after use of the Healthy Minds Innovations app in firefighters. Error bars indicate standard error of the mean. *$P \leq .05$, **$P \leq .01$.
[PNG File , 83 KB - formative_v6i6e34951_app1.png]

References

1. Igboanugo S, Bigelow PL, Mielke JG. Health outcomes of psychosocial stress within firefighters: A systematic review of the research landscape. Jrnl of Occup Health 2021 Mar 29;63(1):e12219-e12219. [doi: 10.1002/1348-9585.12219] [Medline: 33780075]
2. Carleton R, Afifi T, Taillieu T, Turner S, El-Gabalawy R, Sareen J, et al. Anxiety-related psychopathology and chronic pain comorbidity among public safety personnel. J Anxiety Disord 2018 Apr;55:48-55 [FREE Full text] [doi: 10.1016/j.janxdis.2018.03.006] [Medline: 29566981]
3. Firew T, Sano ED, Lee JW, Flores S, Lang K, Salman K, et al. Protecting the front line: a cross-sectional survey analysis of the occupational factors contributing to healthcare workers' infection and psychological distress during the COVID-19 pandemic in the USA. BMJ Open 2020 Oct 21;10(10):e042752 [FREE Full text] [doi: 10.1136/bmjopen-2020-042752] [Medline: 33087382]
4. "Dead on Arrival": A N.Y. Fire Chief's COVID Journal. ProPublica. 2020 Apr 05. URL: https://www.propublica.org/article/dead-on-arrival-a-ny-fire-chiefs-covid-journal [accessed 2020-06-01]
5. Walsh R, Shapiro SL. The meeting of meditative disciplines and western psychology: A mutually enriching dialogue. American Psychologist 2006;61(3):227-239. [doi: 10.1037/0003-066x.61.3.227]
6. Kabat-Zinn J. Full catastrophe living: using the wisdom of your body and mind to face stress, pain, and illness. New York, NY: Delacorte Press; 1990.
7. Smith BW, Ortiz JA, Steffen LE, Tooley EM, Wiggins KT, Yeater EA, et al. Mindfulness is associated with fewer PTSD symptoms, depressive symptoms, physical symptoms, and alcohol problems in urban firefighters. J Consult Clin Psychol 2011 Oct;79(5):613-617. [doi: 10.1037/a0025189] [Medline: 21875175]
8. Haller H, Winkler MM, Klose P, Dobos G, Kümmel S, Cramer H. Mindfulness-based interventions for women with breast cancer: an updated systematic review and meta-analysis. Acta Oncologica 2017 Jul 07;56(12):1665-1676. [doi: 10.1080/0284186x.2017.1342862]
9. Márquez MA, Galiana L, Oliver A, Sansó N. The impact of a mindfulness-based intervention on the quality of life of Spanish national police officers. Health Soc Care Community 2021 Sep 30;29(5):1491-1501. [doi: 10.1111/hsc.13209] [Medline: 33125800]
10. Reangsing C, Punsuwun S, Schneider JK. Effects of mindfulness interventions on depressive symptoms in adolescents: A meta-analysis. Int J Nurs Stud 2021 Mar;115:103848. [doi: 10.1016/j.ijnurstu.2020.103848] [Medline: 33383273]
11. Denkova E, Zanesco AP, Rogers SL, Jha AP. Is resilience trainable? An initial study comparing mindfulness and relaxation training in firefighters. Psychiatry Res 2020 Jan 16;285:112794. [doi: 10.1016/j.psychres.2020.112794] [Medline: 32078885]
12. Kaplan JB, Bergman AL, Christopher M, Bowen S, Hunsinger M. Role of resilience in mindfulness training for first responders. Mindfulness (N Y) 2017 Oct 19;8(5):1373-1380 [FREE Full text] [doi: 10.1007/s12671-017-0713-2] [Medline: 34484459]
13. Joyce S, Shand F, Bryant RA, Lal TJ, Harvey SB. Mindfulness-based resilience training in the workplace: pilot study of the internet-based Resilience@Work (RAW) Mindfulness Program. J Med Internet Res 2018 Sep 11;20(9):e10326 [FREE Full text] [doi: 10.2196/10326] [Medline: 30206055]
14. Counson I, Hosemans D, Lal TJ, Mott B, Harvey SB, Joyce S. Mental health and mindfulness amongst Australian fire fighters. BMC Psychol 2019 Jun 14;7(1):34 [FREE Full text] [doi: 10.1186/s40359-019-0311-2] [Medline: 31200776]
15. Joyce S, Shand F, Lal TJ, Mott B, Bryant RA, Harvey SB. Resilience@Work Mindfulness Program: results from a cluster randomized controlled trial with first responders. J Med Internet Res 2019 Feb 19;21(2):e12894 [FREE Full text] [doi: 10.2196/12894] [Medline: 30777846]
16. Stanley IH, Boffa JW, Tran JK, Schmidt NB, Joiner TE, Vujanovic AA. Posttraumatic stress disorder symptoms and mindfulness facets in relation to suicide risk among firefighters. J Clin Psychol 2019 Apr 23;75(4):696-709 [FREE Full text] [doi: 10.1002/jclp.22748] [Medline: 30672591]
17. Lee J, Lee J, Lee K. Moderated mediation effect of mindfulness on the relationship between muscular skeletal disease, job stress, and turnover among Korean firefighters. Saf Health Work 2020 Jun;11(2):222-227 [FREE Full text] [doi: 10.1016/j.shaw.2020.03.006] [Medline: 32596019]
18. Gonzalez-Hernandez E, Romero R, Campos D, Burychka D, Diego-Pedro R, Baños R, et al. Cognitively-Based Compassion Training (CBCT) in breast cancer survivors: a randomized clinical trial study. Integr Cancer Ther 2018 Sep 22;17(3):684-696 [FREE Full text] [doi: 10.1177/1534735418772095] [Medline: 29681185]
19. Murfield J, Moyle W, O'Donovan A. Mindfulness- and compassion-based interventions for family carers of older adults: A scoping review. Int J Nurs Stud 2021 Apr;116:103495. [doi: 10.1016/j.ijnurstu.2019.103495] [Medline: 31862112]

20. Carson JW, Keefe FJ, Lynch TR, Carson KM, Goli V, Fras AM, et al. Loving-kindness meditation for chronic low back pain: results from a pilot trial. J Holist Nurs 2005 Sep 24;23(3):287-304. [doi: 10.1177/0898010105277651] [Medline: 16049118]
21. Lee J. Perceived social support functions as a resilience in buffering the impact of trauma exposure on PTSD symptoms via intrusive rumination and entrapment in firefighters. PLoS One 2019 Aug 2;14(8):e0220454 [FREE Full text] [doi: 10.1371/journal.pone.0220454] [Medline: 31374563]
22. Lazarus R, Folkman S. Stress, appraisal, and coping. New York, NY: Springer; 1984.
23. Prati G, Pietrantoni L, Cicognani E. Self-efficacy moderates the relationship between stress appraisal and quality of life among rescue workers. Anxiety Stress Coping 2010 Jul;23(4):463-470. [doi: 10.1080/10615800903431699] [Medline: 19937503]
24. Adam EK, Quinn ME, Tavernier R, McQuillan MT, Dahlke KA, Gilbert KE. Diurnal cortisol slopes and mental and physical health outcomes: A systematic review and meta-analysis. Psychoneuroendocrinology 2017 Sep;83:25-41 [FREE Full text] [doi: 10.1016/j.psyneuen.2017.05.018] [Medline: 28578301]
25. Juárez García DM, Landero Hernández R, González Ramírez M, Jaime Bernal L. Variación diurna del cortisol y su relación con estrés, optimismo y estrategias de afrontamiento en mujeres con cáncer de mama. Act Colom Psicol 2016 Apr 10;19(1):103-122. [doi: 10.14718/ACP.2016.19.1.6]
26. Carlson LE, Doll R, Stephen J, Faris P, Tamagawa R, Drysdale E, et al. Randomized controlled trial of mindfulness-based cancer recovery versus supportive expressive group therapy for distressed survivors of breast cancer (MINDSET). JCO 2013 Sep 01;31(25):3119-3126. [doi: 10.1200/jco.2012.47.5210] [Medline: 23918953]
27. Bergen-Cico D, Possemato K, Pigeon W. Reductions in cortisol associated with primary care brief mindfulness program for veterans with PTSD. Med Care. Dec 2014;52(12 Suppl 5):S25-S31. [doi: 10.1097/mlr.0000000000000224] [Medline: 25397819]
28. Koncz A, Demetrovics Z, Takacs ZK. Meditation interventions efficiently reduce cortisol levels of at-risk samples: a meta-analysis. Health Psychol Rev 2021 Mar 07;15(1):56-84. [doi: 10.1080/17437199.2020.1760727] [Medline: 32635830]
29. Healthy Minds Innovations, Inc. Healthy Minds Program. App Store. URL: https://apps.apple.com/us/app/healthy-minds-program/id1326310617 [accessed 2022-05-16]
30. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, PROMIS Cooperative Group. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. J Clin Epidemiol 2010 Nov;63(11):1179-1194 [FREE Full text] [doi: 10.1016/j.jclinepi.2010.04.011] [Medline: 20685078]
31. Malach-Pines A. The Burnout Measure, Short Version. Int J Stress Manag 2005 Feb;12(1):78-88. [doi: 10.1037/1072-5245.12.1.78]
32. Smith TD, Hughes K, DeJoy DM, Dyal M. Assessment of relationships between work stress, work-family conflict, burnout and firefighter safety behavior outcomes. Safety Science 2018 Mar;103:287-292. [doi: 10.1016/j.ssci.2017.12.005]
33. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: The PANAS scales. J Pers Soc Psychol 1988;54(6):1063-1070. [doi: 10.1037/0022-3514.54.6.1063]
34. Zeiders KH, Hoyt LT, Adam EK. Associations between self-reported discrimination and diurnal cortisol rhythms among young adults: The moderating role of racial-ethnic minority status. Psychoneuroendocrinology 2014 Dec;50:280-288 [FREE Full text] [doi: 10.1016/j.psyneuen.2014.08.023] [Medline: 25262035]
35. Zeiders KH, Doane LD, Adam EK. Reciprocal relations between objectively measured sleep patterns and diurnal cortisol rhythms in late adolescence. J Adolesc Health 2011 Jun;48(6):566-571 [FREE Full text] [doi: 10.1016/j.jadohealth.2010.08.012] [Medline: 21575815]
36. Julious SA. Sample size of 12 per group rule of thumb for a pilot study. Pharmaceut Statist 2005 Oct;4(4):287-291. [doi: 10.1002/pst.185]
37. Soteriades E, Smith D, Tsismenakis A, Baur D, Kales S. Cardiovascular disease in US firefighters: a systematic review. Cardiol Rev 2011;19(4):A. [doi: 10.1097/crd.0b013e318215c105] [Medline: 21646874]
38. Brady KT, Killeen TK, Brewerton T, Lucerini S. Comorbidity of psychiatric disorders and posttraumatic stress disorder. J Clin Psychiatry 2000;61(Suppl 7):22-32. [Medline: 10795606]
39. Curhan KB, Sims T, Markus HR, Kitayama S, Karasawa M, Kawakami N, et al. Just how bad negative affect is for your health depends on culture. Psychol Sci 2014 Dec 10;25(12):2277-2280 [FREE Full text] [doi: 10.1177/0956797614543802] [Medline: 25304884]
40. Jahnke SA, Poston WSC, Haddock CK, Jitnarin N. Injury among a population based sample of career firefighters in the central USA. Inj Prev 2013 Dec 16;19(6):393-398. [doi: 10.1136/injuryprev-2012-040662] [Medline: 23504995]
41. Phelps SM, Drew-Nord DC, Neitzel RL, Wallhagen MI, Bates MN, Hong OS. Characteristics and predictors of occupational injury among career firefighters. Workplace Health Saf 2018 Jun 18;66(6):291-301 [FREE Full text] [doi: 10.1177/2165079917740595] [Medline: 29251258]
42. Gordon H, Larivière M. Physical and psychological determinants of injury in Ontario forest firefighters. Occup Med (Lond) 2014 Dec 03;64(8):583-588. [doi: 10.1093/occmed/kqu133] [Medline: 25187490]

43. Kim Y, Ahn Y, Kim K, Yoon J, Roh J. Association between job stress and occupational injuries among Korean firefighters: a nationwide cross-sectional study. BMJ Open 2016 Nov 25;6(11):e012002 [FREE Full text] [doi: 10.1136/bmjopen-2016-012002] [Medline: 27888173]
44. Kendler KS, Hettema JM, Butera F, Gardner CO, Prescott CA. Life event dimensions of loss, humiliation, entrapment, and danger in the prediction of onsets of major depression and generalized anxiety. Arch Gen Psychiatry 2003 Aug 01;60(8):789-796. [doi: 10.1001/archpsyc.60.8.789] [Medline: 12912762]
45. Medvedev ON, Norden PA, Krägeloh CU, Siegert RJ. Investigating unique contributions of dispositional mindfulness facets to depression, anxiety, and stress in general and student populations. Mindfulness 2018 Mar 5;9(6):1757-1767. [doi: 10.1007/s12671-018-0917-0]
46. Hoge EA, Hölzel BK, Marques L, Metcalf CA, Brach N, Lazar SW, et al. Mindfulness and self-compassion in generalized anxiety disorder: examining predictors of disability. Evid Based Complement Alternat Med 2013;2013:576258-576257 [FREE Full text] [doi: 10.1155/2013/576258] [Medline: 24174978]
47. Tomlinson ER, Yousaf O, Vittersø AD, Jones L. Dispositional mindfulness and psychological health: a systematic review. Mindfulness (N Y) 2018 Jul 1;9(1):23-43 [FREE Full text] [doi: 10.1007/s12671-017-0762-6] [Medline: 29387263]
48. Betts KR, Hinsz VB. Mindful attention and awareness predict self-reported food safety practices in the food service industry. Curr Psychol 2014 Jul 22;34(2):191-206. [doi: 10.1007/s12144-014-9251-4]
49. Zhang J, Wu C. Corrigendum to "The influence of dispositional mindfulness on safety behaviors: A dual process perspective" [Accid. Anal. Prev. 70 (2014) 24–32]. Accident Analysis & Prevention 2015 Mar;76:166. [doi: 10.1016/j.aap.2014.12.003]
50. Dierynck B, Leroy H, Savage GT, Choi E. The role of individual and collective mindfulness in promoting occupational safety in health care. Med Care Res Rev 2017 Feb 03;74(1):79-96. [doi: 10.1177/1077558716629040] [Medline: 26825941]

Abbreviations

AUC: area under the curve
CAR: cortisol awakening response
HMI: Healthy Minds Innovations
HPA: hypothalamic-pituitary adrenal
PANAS-SF: Positive and Negative Affect Schedule
PROMIS: Patient Reported Outcomes Measurement Information System
PTSD: posttraumatic stress disorder

Edited by A Mavragani; submitted 14.11.21; peer-reviewed by HJ Jeon, CF Lai; comments to author 04.03.22; revised version received 15.04.22; accepted 18.04.22; published 08.06.22.

Please cite as:
Pace TWW, Zeiders KH, Cook SH, Sarsar ED, Hoyt LT, Mirin NL, Wood EP, Tatar R, Davidson RJ
Feasibility, Acceptability, and Preliminary Efficacy of an App-Based Meditation Intervention to Decrease Firefighter Psychological Distress and Burnout: A One-Group Pilot Study
JMIR Form Res 2022;6(6):e34951
URL: https://formative.jmir.org/2022/6/e34951
doi:10.2196/34951
PMID:35675115

Original Paper

Toward Improved Treatment and Empowerment of Individuals With Parkinson Disease: Design and Evaluation of an Internet of Things System

Liran Karni[1], MSc; Ilir Jusufi[2], PhD; Dag Nyholm[3], MD, PhD; Gunnar Oskar Klein[1], MD, PhD; Mevludin Memedi[1], PhD

[1]Centre for Empirical Research on Information Systems, Örebro University School of Business, Örebro, Sweden

[2]Department of Computer Science and Media Technology, Linnaeus University, Växjö, Sweden

[3]Department of Medical Sciences, Neurology, Uppsala University, Uppsala, Sweden

Corresponding Author:
Liran Karni, MSc
Centre for Empirical Research on Information Systems
Örebro University School of Business
Fakultetsgatan 1
Örebro, 702 81
Sweden
Phone: 46 19 301218
Email: liran.karni@oru.se

Abstract

Background: Parkinson disease (PD) is a chronic degenerative disorder that causes progressive neurological deterioration with profound effects on the affected individual's quality of life. Therefore, there is an urgent need to improve patient empowerment and clinical decision support in PD care. Home-based disease monitoring is an emerging information technology with the potential to transform the care of patients with chronic illnesses. Its acceptance and role in PD care need to be elucidated both among patients and caregivers.

Objective: Our main objective was to develop a novel home-based monitoring system (named EMPARK) with patient and clinician interface to improve patient empowerment and clinical care in PD.

Methods: We used elements of design science research and user-centered design for requirement elicitation and subsequent information and communications technology (ICT) development. Functionalities of the interfaces were the subject of user-centric multistep evaluation complemented by semantic analysis of the recorded end-user reactions. The ICT structure of EMPARK was evaluated using the ICT for patient empowerment model.

Results: Software and hardware system architecture for the collection and calculation of relevant parameters of disease management via home monitoring were established. Here, we describe the patient interface and the functional characteristics and evaluation of a novel clinician interface. In accordance with our previous findings with regard to the patient interface, our current results indicate an overall high utility and user acceptance of the clinician interface. Special characteristics of EMPARK in key areas of interest emerged from end-user evaluations, with clear potential for future system development and deployment in daily clinical practice. Evaluation through the principles of ICT for patient empowerment model, along with prior findings from patient interface evaluation, suggests that EMPARK has the potential to empower patients with PD.

Conclusions: The EMPARK system is a novel home monitoring system for providing patients with PD and the care team with feedback on longitudinal disease activities. User-centric development and evaluation of the system indicated high user acceptance and usability. The EMPARK infrastructure would empower patients and could be used for future applications in daily care and research.

(JMIR Form Res 2022;6(6):e31485) doi:10.2196/31485

KEYWORDS

Internet of Things; wearable technology; Parkinson disease; patient empowerment; objective measures; self-assessment; self-management; web interface

Introduction

Parkinson disease (PD) is a chronic degenerative disorder that causes progressive neurological deterioration with profound effects on the affected individual's quality of life [1-4]. It affects 1 out of 100 people aged >60 years, and its prevalence has doubled since 2005 [5]. The progression of the movement disorder is often accompanied by the development of neuropsychiatric deficits in the later stages. The characteristic fluctuations of various neurological symptoms become increasingly pronounced over the years [6]. Clinicians and patients find it challenging to establish a treatment strategy based on the patient's ability to plan and self-administer levodopa to temporarily counteract the motor deficits. Over time, it often becomes difficult to stabilize the fluctuations in the neurological symptoms, and patients may need more advanced treatments based on specialist assessments [7]. Current disease management strategies rely on scheduled in-hospital assessments and periodic contacts with expert physicians. However, in Europe, 44% of newly diagnosed people with PD do not meet neurologists with an interest and expertise in PD despite recommendations to do so [1]. Patients also have limited access to specialist nurses and occupational therapists owing to resource limitations or patient's disabilities [5,8,9]. In addition, periodical snapshot assessments performed during in-hospital visits carry the possibility of over- or underestimation of the longitudinal disease burden and cannot reliably describe the situation these patients face in their home environment [4,5,9-11].

Therefore, there is a substantial interest in developing home-based care models for PD [5,12]. The main goal of this approach is to provide autonomy and self-efficacy for self-management through an increased understanding and interpretation of the associations among disease symptoms, daily life activities, and treatment [13,14]. Technological advancements can largely contribute to improvements in the home-based care of people with PD [4,12,15-18]. For instance, several device-aided treatments are available for managing motor fluctuations in people with PD, such as deep brain stimulation, infusions of apomorphine or levodopa [19,20] and microtablets of levodopa [21]. Such advanced treatments may be tailored to the needs of each individual. The Task Force on Technology of the International Parkinson and Movement Disorder Society has called for tighter collaboration among clinicians, researchers, patients, and engineers to achieve this goal by developing home-based integrated measurement and closed-loop therapeutic systems with high patient adherence [12]. The task force also envisioned that this strategy could lead to the adoption of clinical-pathophysiological phenotyping and early detection of milestones in disease progression, improve subgroup targeting of patients for future testing of disease-modifying treatments, and identify objective biomarkers to improve longitudinal tracking of impairments in clinical care and research [12].

The possibility of home-based continuous monitoring of PD symptoms is now possible using new technologies available today, such as the Internet of Things (IoT), in the form of wearable sensors to capture fluctuating disease activities [16,22], potentially serious incidents such as falls or freezing of gait [23-29], or assistance with the medication strategy [30]. There is also an expectation that the technology-based objective measurements may, in the future, replace or amend the current practice of disease severity assessment and follow-up that mostly rely on patient-reported outcomes and clinical scales [14,30,31]. For instance, in our previous publications, we demonstrated good validity, reliability, and responsiveness to treatment changes using objective measures based on multiple wearable sensors for motor function assessment [32,33].

Despite these technological advances, it is still unclear how these tools could be translated into a better patient self-management and mutual care plan or how they could improve the overall health outcomes [34]. Some important questions that remain to be answered include what clinical or patient needs should be addressed, by which technology, and how these systems could be deployed in patients' homes to maximize treatment benefits and compliance, while minimizing the risk of disruption to daily life [11,12,14]. Based on these considerations, the Movement Disorder Society Task Force on Technology has recently deliberated that future research and development should be based on the identification of clinically meaningful targets for modification instead of a pure technology-driven selection of measurement end points [12].

Targeting patients by providing feedback about their disease activities and their related self-management is a promising strategy for PD. Prior publications almost exclusively focused on patient home assessment and measurements as tools for clinical decision-making and targeted clinicians as end users. Home monitoring of disease activities can increase the extent and quality of self-engagement, which in turn can potentially increase patient satisfaction and lead to positive changes in the health-related quality of life and outcomes [12,13,31,35]. This symptom-driven self-management strategy can be especially successful for people with PD who are knowledgeable about the disease functions [36]. It is presumable that such patient-targeted feedback efforts would empower people with PD with noticeable differences in the efficacy of self-management, patients' communication with the care team, and improved health-related outcomes [36].

We recently published the outlines of an IoT-based system (named EMPARK) aimed at empowering people with PD and their care team [32,33]. The system is based on home monitoring, with the aim of collecting longitudinal data for patient self-assessment and clinical decision-making. We identified patient empowerment as an important primary end point of information and communications technology (ICT) interventions in people with PD. In a subsequent publication, we described the design of a patient interface prototype for EMPARK along with its initial assessment with people with PD [32,33]. The interface was designed to visualize symptom and medication information collected by an IoT-based system comprising a mobile technology (tablet), electronic dosing device, wrist sensor, and bed sensor. The results indicated that patients found visual feedback on motor function, sleep, medication compliance, meal intake timing, and their relationship to self-medication as a useful tool that could

improve their daily self-management and mutual care planning with their physicians.

Improving patient empowerment has been a key concept behind the development of EMPARK. A retrospective evaluation of its ICT design against the ICT for patient empowerment model (ICT4PEM) [37], together with the results of the patient interface evaluation, indicated that EMPARK has a strong potential to empower its primary users, the patients [32,33].

Although the EMPARK system has not yet been deployed in practice, in this study we demonstrate its system architecture, implementation, and evaluation of user interfaces. As EMPARK is a project under development, our demonstration builds on our prior results regarding the data and interface architecture of the system [32,33]. As part of this demonstration, we also provide evidence of how the EMPARK system could empower patients and members of the care team.

Methods

Overall Study Design and Theory

The EMPARK system was developed based on the principles of design science research [38], user-centered design [39,40], and the overall target of patient empowerment. As an initial step, a set of long-term goals was agreed upon by the authors and through iterative consultation with clinical experts in PD. These goals also served as the foundation for the development of the overall and specific requirements for system design. These requirements were regarded as the core tasks of the EMPARK system development.

After a thorough study of the literature and several iterations with experts, we developed the ICT4PEM, a novel framework to support future ICT interventions that primarily aim to improve patient empowerment [37]. To the best of our knowledge, the ICT4PEM is the only framework that provides clear, International Organization for Standardization–style definitions for core empowerment characteristics (control, psychological coping, self-efficacy, understanding, legitimacy, and support) and allocates them with project-specific, well-defined targets for ICT interventions. This framework defines the core characteristics of the empowerment concept as the primary target of ICT interventions. It also defines a set of empowerment consequences as potential targets of indirect intervention, such as expressed patient perceptions, behavior, clinical outcomes, and health system effects. ICT4PEM stratifies interventional design into strategies, narrowed further into ICT services, and finally, specific ICT tools. To ensure the goals of empowerment, targeting ICT4PEM requires that each ICT service within a project must be allocated to specific empowerment domains with strict definitions. Finally, ICT4PEM is intended to be used for project evaluations with the aim of demonstrating the effects on patient empowerment or its consequences [37].

We applied the ICT4PEM to evaluate the ICT design of the EMPARK system in terms of its relevance and strength for patient empowerment. This was possible by identifying well-defined patient empowerment characteristics within the ICT4PEM that could be easily identified as the primary targets of the EMPARK system interventions. Consequently, we categorized the ICT interventions of EMPARK by strategies, as described in ICT4PEM, and allocated these strategies and ICT interventions to the relevant empowerment characteristics. This analysis of the EMPARK design with the ICT4PEM framework was based on our results from an extensive evaluation of the patient mock-up interface [32,33] by patient users. In addition, evaluation of the clinician interface revealed additional feedback on the patient-empowering potential of EMPARK from clinician users.

Study Variables

Selection of Study Variables

PD is a complex and multimodal neurological disorder that affects the patients' quality of life. Therefore, the system collects multidimensional data to represent the overall health status of the patients [12,40]. The concept of overall health status, based on consultations with participating expert neurologists and patients, was defined as a data set that incorporates both objective and self-reported measures of ongoing disease activities and, regarding the disease-specific symptoms, includes both motor and nonmotor functional parameters.

Furthermore, decisions regarding the inclusion of variables were made by examining the set of requirements shown in Textbox 1.

The final selection of individual variables and relevant technologies for home implementation and data acquisition was based on prior literature and recommendations from the Movement Disorders Society Task Force [12] and by consultation with expert physicians and patients. Subsequently, this selection was refined based on the evaluation of mock-up interfaces for patients and physicians.

Textbox 1. Requirements.

- Potentially empower the patients if they receive feedback on it.
- Improve understanding of their disease and its relationship with their daily activities and self-management decisions.
- Improve self-efficacy and control of daily disease management.
- Improve the overall feeling of support in their self-management and communication with the care providers.
- People with Parkinson disease, except those with very advanced disease, are capable of self-reporting (regarding self-reported data only).
- Availability of reliable technology for its objective measurement, and that the technology is nonintrusive in the home environment and nondisrupting regarding the patient's regular daily activities.
- Capable of stable wireless data transfer.
- Preferentially sensor technology with the capacity to provide data for multiple end points (core data).
- The least possible maintenance dependent.
- Data are assessed as clinically relevant by expert physicians.
- Data are assessed as scientifically relevant for future studies regarding their clinical validity.

Calculation of Study Variables

Overview

The EMPARK system collects health information from the home environment of patients. The system summarizes the objective data (medication compliance, sleep and motor function) as well as subjective data (meal intake time, physical exercise and quality of life) and calculates the corresponding scores per calendar day. All the scores were produced in the range of 0 (low) to 100 (high). Most of these parameters have been previously described elsewhere [32]. Nevertheless, we provide a list of these parameters with their definitions and a brief description of the method of calculation.

Medication Compliance Score

Using the planned and delivered medication intake times, the medication compliance score was calculated by first calculating the deviation (in minutes) as depicted in the following equation:

where $Planned_i$ is the time of a planned medication intake and $Taken_i$ is the time of a delivered medication intake. This resulted in scores ranging from 0 to 100. The final medication compliance score per day was calculated using the following equation:

where n is the total number of medication intake occasions per day. Patients can take an extra dose to either replace an intended dose that was initially missed or to take another dose. Extra doses administered within 30 minutes before or after a planned time were considered as replacement and omitted in calculations, whereas, for other extra doses, the compliance score was set to the highest score, that is, 100.

Sleep Score

This was obtained by using the time information recorded by the bed sensor. A sleep score was calculated based on 3 individual components: sleep duration, habitual sleep efficiency, and sleep disturbance. The components were derived based on the Pittsburgh Sleep Quality Index and the work of Williams and Cook [41].

To calculate sleep duration, the system first calculated the differences (in total hours) between the time when the bed was occupied and when it was empty and calculated the cumulative hours being in bed:

where n is the total number of recorded intervals when the individual was in bed. The following rules were applied to calculate sleep duration:

- *IF Cumulative Hours Being In Bed* > 7 *THEN Sleep Duration* = 0
- *IF Cumulative Hours Being In Bed* > 6 *AND Cumulative Hours Being In Bed* < 7 *THEN Sleep Duration* = 1
- *IF Cumulative Hours Being In Bed* > 5 *AND Cumulative Hours Being In Bed* < 6 *THEN Sleep Duration* = 2
- *IF Cumulative Hours Being In Bed* < 5 *THEN Sleep Duration* = 3

The second component was the habitual sleep efficiency score, which was calculated by first calculating the ratio between total hours being in bed with interruptions and total hours being in bed without interruptions using the following equation:

where *Min Time* is the time when the patient went to bed for the first time during a particular day and *Max Time* is the time when the patient left the bed the next day. The habitual sleep efficiency score was calculated according to the following rules:

- *IF Ratio* ≥ 85 *THEN Habitual Sleep Efficiency* = 0
- *IF Ratio* > 75 *AND Ratio* <85 *THEN Habitual Sleep Efficiency* = 1
- *IF Ratio* > 65 *AND Ratio* <75 *THEN Habitual Sleep Efficiency* = 2
- *IF Ratio* < 65 *THEN Habitual Sleep Efficiency* = 3

The third component that was calculated was sleep disturbances score using the following steps:

Step 1: Count the number of interruptions during sleep, which resulted in *Total Number Of Interruptions* score.

Step 2: Apply the following rules:

- *IF Total Number Of Interruptions* = 0 *THEN Sleep Disturbances* = 0
- *IF Total Number Of Interruptions* = 1 *THEN Sleep Disturbances* = 1
- *IF Total Number Of Interruptions* = 2 *THEN Sleep Disturbances* = 2
- *IF Total Number Of Interruptions* ≥ 3 *THEN Sleep Disturbances* = 3

Step 3: Calculate the final sleep score using the following equation:

Motor Function Score

The wrist sensor-based system produced scores every 2 minutes for bradykinesia and dyskinesia. To summarize bradykinesia and dyskinesia scores per day, initially the mean values were calculated.

Meal Timing Compliance Score

An algorithm calculated the alterations from the recommended meal intake time (30 minutes before and 60 minutes after levodopa medication intake) to calculate the score:

where *NWI* is the number of reported occasions when meal intake time was within the recommended interval of 30 minutes before and 60 minutes after medication intake, and *n* is the total number of reported occasions of meal intake during a day.

Physical Activity Score

The score per day was calculated using the following equation:

where $Exercise_{duration}$ is in minutes, $Exercise_{mode}$ is the reported physical exercise mode (1: boxing, dancing, running, and swimming; 2: bicycling and gym; and 3: walking), $Patient_{target}$ is the individual patient target ranging from 0 to 100, and *n* is the total number of physical exercise occasions reported per day. As there may be more than one exercise occasion, the algorithm first calculates an average score per day followed by its normalization using the patient target. For instance, if a patient reported 2 exercise occasions during a particular day with a target of 80 with the following data: for the first occasion the patient reported exercise mode of 1 and duration of 10 minutes, for the second occasion the patient reported exercise mode of 2 and duration of 25 minutes then the physical exercise score for the day would be 75 × {[10 + (2×25)]/80×100}.

Quality of Life Scores

This was derived from each day the patients use a custom application to answer 9 health-related questions (Table 1) taken from the 8-item Parkinson Disease Questionnaire [42] and the EQ-5D-3L (European Quality of Life 5 Dimensions 3 Level) health questionnaire [43]. Inclusion or exclusion of the questions from the 8-item Parkinson Disease Questionnaire and EQ-5D-3L was performed in collaboration with clinicians at Uppsala and Örebro University Hospitals. The first 8 questions are scored on 3 levels: no problems (level 1), some problems (level 2), and extreme problems (level 3). The last question is about the overall health when the patients are asked to rate their health on a visual analog scale ranging from 0 to 100 where 0 is the *worst imaginable health* and 100 is the *best imaginable health*.

Table 1. Quality of life items and source questionnaires.

Item number	Item	Questionnaire
1	Mobility	EQ-5D-3L[a]
2	Personal care	EQ-5D-3L
3	Daily activities	EQ-5D-3L
4	Pain or discomfort	EQ-5D-3L
5	Worry or depression	EQ-5D-3L
6	Concentration difficulties	PDQ-8[b]
7	Communication difficulties	PDQ-8
8	Painful cramps or spasms in the muscles	PDQ-8
9	Overall health	EQ-5D-3L

[a]EQ-5D-3L: European Quality of Life 5 Dimensions 3 Level.

[b]PDQ-8: 8-item Parkinson Disease Questionnaire.

Elicitation of the User-Specific Requirements

Requirements related to patient and clinician interfaces were gathered through discussions with the end users based on the principles of user-centered design [39].

The development of these requirements for the patient interface has been discussed elsewhere [32,33]. This section provides an outline. The patient interface had 4-stage patient- and expert-based development cycles that revealed 8 categories of information to be presented to the patients.

Regarding the clinician interface, a broad set of requirements was identified during 5 group discussions or brainstorming design [44] sessions between the authors, medical specialists in neurology with strong clinical backgrounds and substantial experience with the management of people with PD, and nurses from the neurology departments at the university hospitals of Uppsala and Örebro, Sweden. During the first session, medical doctors and nurses were informed about the overall system architecture of the proposed EMPARK system. During this session, the participants outlined the requirements that they anticipated the clinician interface would offer them. These requirements were documented by taking notes and later elaborated upon by the design team. The design principles of the clinician interface were derived from these sessions.

Evaluation of the Users' Interfaces

The mock-up patient interface was subject to extensive user-centric evaluation as previously published [32]. We provide a summary of the evaluation process. We created simulated patient scenarios to be used during the development and evaluation of the patient interface. Each patient case was based on a database developed by clinical domain experts. The database was in an Excel file format, which was subsequently loaded into the server, and presented to the patients as test subjects during the late development and evaluation phases of the patient interface as published earlier. These mock-up scenarios were later used to determine user acceptance [33]. After heuristic evaluation of the initial patient interface, subsequent evaluation was performed by requesting patients to perform tasks on the mock-up interface that mimicked daily life tasks based on realistic patient data. The patients needed to use the capacities of the interface to summarize disease activities by week or day, alternatively showing data at a higher temporal resolution.

Similarly, simulated clinical scenarios of PD with corresponding data sets were developed by collaborating expert physicians to evaluate the clinician interface. The evaluation was performed with the help of 2 professionals (a neurologist and nurse) from the Uppsala University Hospital. After a brief introduction to the functionalities of the interface, the users had to complete different tasks using the prototype clinical interface representing the simulated clinical scenario, following a semistructured interview. The participant were instructed to comment loudly on their thoughts during the completion of the tasks according to the *think aloud* methodology [45]. The task-based evaluation session was documented by the investigators during a one-to-one session with the evaluator and was later used for further refinement of the prototype clinician interface. This first task-based evaluation step intended to assess the overall usability of the interactive functions and visual solutions of the interface and included simple instructions. In the subsequent step, the participants were asked to complete 3 complex tasks related to specific clinical case scenarios to test the functionalities of the prototype and the clinicians' understanding. This was video recorded using a screen-capture software for later documentation and analysis (Multimedia Appendix 1).

We also conducted semistructured interviews (Multimedia Appendix 2) with the same 2 participants during the same session. The development of questions and evaluation of the responses that were part of the semistructured interviews followed the methodology heuristic evaluation and questions [46] and the Computer Usability Questionnaire [47]. The questions were developed through iterative brainstorming meetings among the researchers. The responses were audio recorded for subsequent transcription and analysis.

Ethics Approval

Ethics approval from the ethics review authority will be applied for at the clinical trial stage according to Swedish law. However, this paper communicates the early evaluation of the clinical interface that was using dummy data with no personal information. This interpretation of the lack of requirement for ethics approval for this step has been checked with the internal policy of our universities.

Results

Requirement Elicitations

In accordance with the overall study design, the first stage of EMPARK development was the determination of its major goals. The list and definitions of these goals are as follows:

- System development: to develop an IoT system comprising sensors and mobile technology to deliver home monitoring of objective motor function, medication use, and patient-reported symptoms and outcomes.
- Patient empowerment: to empower patients and improve their self-management through a better understanding of their disease with the help of feedback from the system.
- Clinician support: to provide physicians with relevant and useful information derived from the system regarding symptoms, treatment response, and individual patient coping strategies for better clinical assessment and treatment strategy.
- Improvement and enablement of research: to establish the EMPARK system as a tool for researchers to better understand the complexity of the disease and to develop and monitor new drugs.

The core tasks of the EMPARK system development were defined as described in the *Methods* section to accomplish the major goals. These included selection of patient-related data for longitudinal acquisition, development of visual interfaces to facilitate follow-up of the health status of people with PD by clinicians, as well as to provide feedback to patients on their data.

User Interface Requirements

The mock-up patient interface was functional at the time of the development of the clinician interface. The requirements elicited for the patient interface were also found to be satisfactory by the participating medical professionals. These requirements have been published previously [32]. Briefly, these included motor function, sleep, medication compliance, meal intake times, physical exercise, and a subjective assessment of the day as parameters. Patient requirements for data representation included routines of an ideal day, effect of meal timing on medication, single day parameter overview, and revealing the effects of motor functions on various other parameters and daily routines. Detailed and aggregated data were table-based, with the possibility of narrowing down the selection.

The clinician interface aims to provide access to aggregated and analyzed data of clinical interest. Furthermore, the patient interface can be shared with the physician or care team for shared decision-making and elaboration of a mutual care plan. The main objectives behind the development of a clinician interface were the elaboration of a web-based application for the analysis and representation of EMPARK data. These objectives were defined based on the following list of specific requirements elicited by iterative consultations with the care team:

- Provide the care team with data on longitudinal fluctuations of disease symptoms, daily patient activities, and self-medication.
- Help clinicians discover disturbances, limitations of disease self-management, or aberrations from the care plan.
- Facilitate communication and co–decision-making with patients about treatment based on objective and longitudinal data on disease activities and self-management.
- Help the care team foster patient empowerment based on objective data.

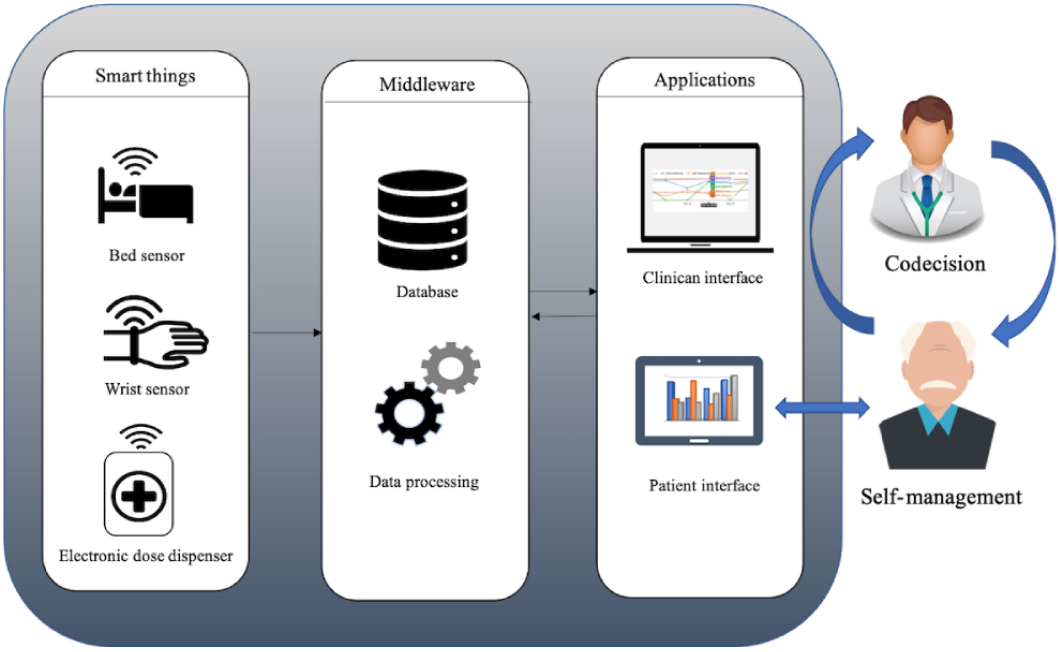
Physicians and nurses also requested access to and visualization of raw data in addition to integrated summaries. A further requirement was an overview graph representing the daily summary with scores from different variables, such as medication, mealtime, exercise, self-assessment, sleep score, and the 2 motor function scores, such as those for bradykinesia and dyskinesia. Other requests were a flexible and interchangeable time axis for the score results with temporal zoom function, allowing discoveries of temporal associations between different disease activities and treatment and the possibility of selecting variables for representation.

EMPARK System Architecture

Overview

The EMPARK system (Figure 1), which is currently under development, leverages IoT technologies to empower people with PD and assist clinicians in making informed decisions. The system is deployed at the patient's home and consists of a set of commercial IoT devices: a wrist sensor that provides measures of bradykinesia and dyskinesias; a medication dosing device that dispenses individualized doses of levodopa based on schedules defined by clinicians; a bed sensor attached to the bed frame that captures data about sleep patterns (time and quality); and a tablet app to collect self-reported data on physical activity, health-related questions, and meal intake time. Data were collected continuously 24×7 and fed into a database. The data processing layer aggregates and analyzes the provided data. A detailed description of the architecture is provided in our previous study [32]. The system has 2 interfaces: a tablet-based app for the patient and a web-based interface for the clinician. Although the data foundation for both interfaces is identical, the use and presentation of the data are tailored to the user's needs.

Figure 1. EMPARK architecture.



The functional EMPARK system has 2 intercalated application cycles: one for patient empowerment and self-management and one for codecision and facilitation of mutual care plan development (Figure 1). The self-management cycle represents home-based disease monitoring and direct feedback of information to the patient through the patient interface. Self-management becomes cyclic when the feedback provided modifies patient self-management, which consequently alters the results of home monitoring. The codecision cycle represents cyclic changes in patient self-management based on the patient's interaction with the care team. As the source of the clinical decision-making is home monitoring, the latter is positioned in the middle of the 2 circles.

Hardware and Software Specifications

The data processing layer, which consists of algorithms for processing data from different sensors and a custom software for summarizing the scores, is written in the C# programming language. The web application (clinician interface) was developed in Python 3.6 using Dash Plotly for visualization. The tablet app (patient interface) was written in C# using Microsoft Xamarin for Android, and SQLite was used to access the data. The principles of the Model-View-View Model Prism architectural design pattern were followed for the development of the patient interface app to ensure high-level cohesion.

User-Specific Functionalities

Patient Interface Functionalities

This section is based on our 2 previous publications on the EMPARK patient interface. The first publication describes user-centered design and evaluation processes and provides a detailed description of interface functionalities [32]. The second publication elaborates the factors that influence user acceptance [33]. As part of the ongoing development of EMPARK, and based on the results of the evaluation, some changes have been implemented in the patient interface. These changes were intended to improve the overall functionality of the interface and user acceptance. Briefly, the previously published version of the patient interface allowed users to select from 3 major representation modes. The start page showed color-coded icons representing the mean values for the last 14 days of the 6 main parameters: overall day score, medicine score, motor function score, physical exercise, meal, and sleep scores. Good, average, and bad scores were defined as >70, 30 to 70, or <30 in value and color coded as green, gray, or orange, respectively. Patients could navigate by clicking on a parameter leading to a 3-week summary of the selected score, showing the daily, color-coded results. From here, patients could select a specific week to see each day of the week as a horizontal graph with graphical representation of the sleeping hours, times of meals and exercises, and the overall daily score in color coding. Patients could also select a day view with a similar timeline-based visualization of the scores (with turn on and off for the presentation of each variable). This page allowed the user to identify temporal associations among elements of self-management (ie, medicine intake or mealtime) with outcome parameters, such as motor function score or sleep quality.

In the revised version of the patient interface, the following changes have been made:

1. The navigation from the home screen was improved by new functional and design elements that allowed the user to select and open a 2-week view, a 1-week view, and a 1-day view, as well as open views for the 6 scores according to their actual preference (Figure 2).
2. The 1-day and 1-week views and the motor function (bradykinesia and dyskinesia) scores are presented separately, in contrast to the prior presentation of their averaged values (Figure 3).
3. Similarly, the temporal associations among the 6 different scores (medication compliance, sleep, bradykinesia, dyskinesia, meal compliance, and physical activity over the last 2-week period) became easier to realize, as we completed this view with additional bar and line graphs (Figure 4).
4. Finally, to facilitate the discovery of correlation among the relevant parameters, the correlation view was modified. It is now possible to change the y-axis to the 2 motor function scores (Figure 5).

Figure 2. Navigation (home page; translated from Swedish to English).

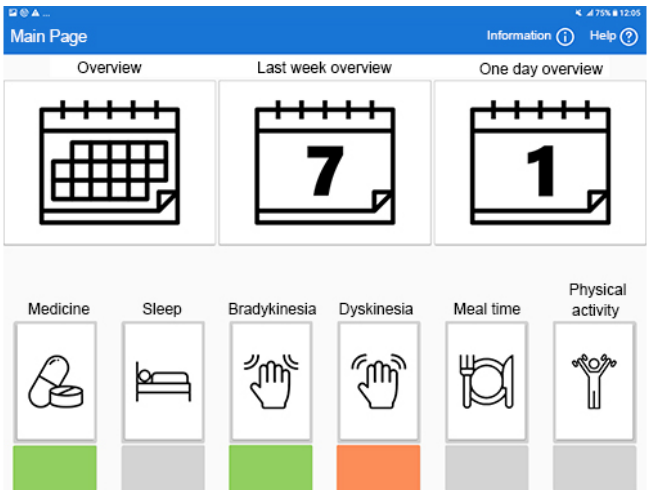


Figure 3. One-day view with separated bradykinesia and dyskinesia motor function scores (translated from Swedish to English).

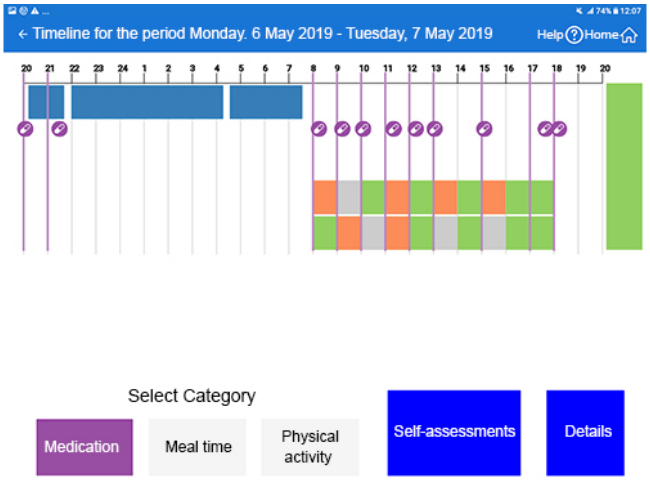


Figure 4. Bar or line graph (translated from Swedish to English).

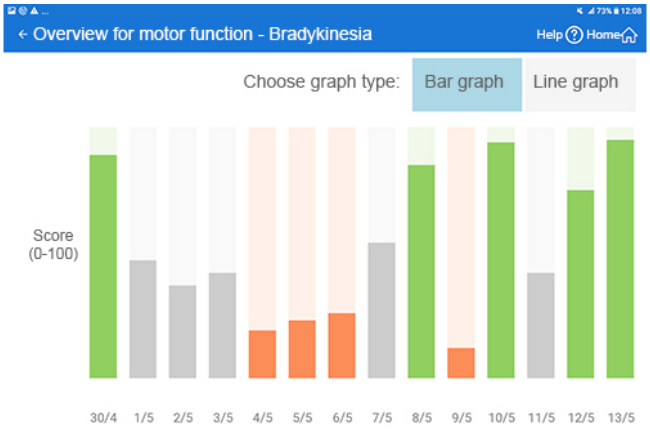
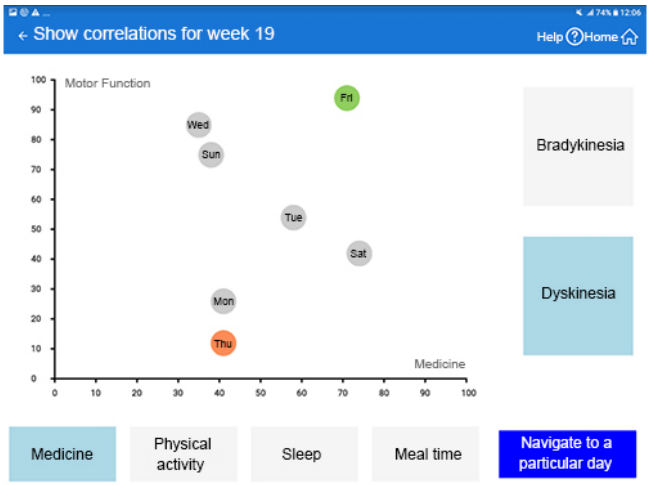


Figure 5. New correlation view of motor function score versus medication compliance score (translated from Swedish to English).



Clinician Interface Functionalities

The development of the clinician interface application was based on specific user requirements as previously described.

To enable rapid patient status assessment, information in the clinician interface is displayed and ordered using a top-down approach with a general overview of the patient's scores and assessments. The main page displays the patient's Daily Summary graph (Figure 6). The Daily Summary is a collection of graphs of the selected variables that can be updated interchangeably: medication score, mealtime score, exercise score, self-assessment, sleep score, and the 2 motor function scores (bradykinesia and dyskinesia). The horizontal axis displays the date, and the vertical axis displays the patient's score on a scale ranging from 0 to 100, where 0 is considered

as the worst score and 100 is regarded as the best. The user can add or remove variables by selecting or deselecting them.

The raw data and more detailed information for each component are shown in speared graphs, allowing the user to subsequently switch between components by using the checkbox list: (1) medicine intake compliance, (2) self-reported meal intake times, (3) exercise, (4) self-reported details, (5) sleep details, and (6) movement indicators.

Some of the main components are the composites of several variables. Self-reported details are composed of several subjective variables (cramps or spasms, communication, worry or depression, pain or discomfort, daily activity, personal care, and mobility) that follow a simple scoring system (0-2), where 0 represents the worst result and 2 represents the best possible result. For the most comprehensive overview of all these parameters along the time scale, we chose a color-coded tile for representation. Sleep details represent sleep duration, efficiency, and disturbances in separate graphs. Movement indicators show bradykinesia and dyskinesia as variables. Following the requirements of the end users, these graphs represent the time plot of the raw wrist sensor data, which neurologists are already familiar with.

We used standard time-series data visualization techniques [48]. Some facets were changed slightly to accommodate specific tasks. The most prominent is the medicine intake compliance chart. The grayed-out line plot shows a cumulative dosage taken per day. The gray circles with horizontal lines represent the planned medicine intake, whereas its vertical positioning corresponds to the planned dosage value. The purple circles indicate the actual dosage taken for the planned intervals. Patients sometimes receive extra doses as indicated by the red circles.

Clicking on any designated point over the graph prompts a pop-up window that displays the exact values for that point of time. Moreover, it is possible to zoom in and out by selecting a section of interest. This action will affect all graphs simultaneously.

Figure 6. A snapshot of the clinician interface. The first graph shows a summary of the scores per day during the last 2 weeks. In the detailed views, users can select or deselect variables: medicine intake compliance, self-reported meal intake times, sleep details, movement indicators, exercise, and self-reported details as a complement to the first graph.

Results From User Interface Evaluation

The patient interface, as previously published, was subject to extensive evaluation [32,33]. Overall, the patients found the visualizations clear and easy to understand and could successfully perform the tasks.

Here, we demonstrate the results of clinician interface evaluation.

The first task series (Multimedia Appendix 1) included instructions to assess the direct usability of the interface functionalities. The users were introduced to the interface shortly before completing these tasks but without specific descriptions and demonstrations of all user functionalities in a visualization of the clinical parameters. Hence, the completion of the tasks was largely based on the intuitive application of the interface. Both users accomplished these tasks with only minor delays and without the need for an intervention from the evaluator. The users also assessed the functionalities as easy to comprehend and in accordance with the expected level of information technology literacy and experiences of an average clinician or nurse user.

The second task series (Multimedia Appendix 2) aimed to evaluate the functionality of the interface in solving complex clinical questions using task-based evaluation. Specifically, the tasks were developed to be a logical continuation of a problem-solving process in which the users were subject to discovering points of interest in the clinical case scenarios.

Notably, there was a marked difference in the response times between the physician and nurse evaluators. The former

completed each task faster by using the functionalities of the interface in a more intuitive way. Similarly, we observed an easier comprehension of the complexity of the represented clinical data by physicians compared with the nurses. The physician completed these tasks with only minimal loud comments that did not reveal any experienced difficulties.

In contrast, the nurses required a substantially longer time to solve the same tasks. The delay in task completion by the nurse evaluator was also accompanied and well explained by the loud comments about the difficulties. In agreement with our own observations, these comments suggested that the nurses experienced more difficulties in comprehending the complex clinical data.

The observations from the analysis of the task completion performance were further elucidated through subsequent semistructured interviews (Multimedia Appendix 3). As the answers for one specific question often included information that corresponded to the other questions, we recategorized the responses by analyzing each sentence separately. To increase the validity of the responses, we analyzed the transcripts for eventual contradictions; for instance, contradictory responses from the same interview participant at different time points in the interview. However, we were unable to identify these statements.

The overall combined assessment of the interview transcripts and observations made by task completion led to the identification of key areas of user interest. These areas were also identified as the main pathways for future development of the system.

We summarized our findings by demonstrating relevant sentences from the evaluators for each defined area (Table 2). The responses from the interview participants were categorized by topic and content analyses. Incongruence among the interview participants is not shown separately but is instead highlighted by the coexistence of both satisfactory and critical comments regarding the same area. Specifically, our nurse participant was overall more critical of the interface dashboard, and these critical comments were mostly about the design of the interface.

Table 2. Summary of user evaluation results for the clinician interface.

	Benefits	Suggested development or modification
Intuitiveness	Enables clinical staff by providing otherwise-unavailable clinical data of great potential relevance; empowers physician-patient relationship; enables potential research in data analysis at large.	Automated correlation analysis among different parameters of the system.
Design	Simple overview and organization of information. Easy to accomplish most tasks at once (physician's view).	Improvement of overall design; time-scale harmonization among parameters and zooming function; making data more clearly understandable.
Empowering the health care provider and usefulness in clinical practice	Novel and useful information content; objective clinical decision-making; comparison with patient's subjective experiences; can improve codecision making; allows early distant identification of patient with nonadherence or unsatisfactory disease control; identification of area for intervention in management; research possibilities by analyzing EMPARK data at a population level.	Motor function (wrist sensors) data presentation to be harmonized with available commercial systems (eg, Parkinson KinetiGraph) [49]; relevant treatments beside levodopa; further patient-derived information to explain fluctuations of parameters; same visual platform (ie, patient interface) for codecision making.
Empowering the patient	Can improve codecision making.	Patient messaging (also to empower the provider).

Both users appreciated the novelty and intuitiveness of the EMPARK system regarding its potential for making important disease-relevant data readily available to the clinical practice. Similarly, both users highlighted that in their routine clinical practice, they lack information that the EMPARK system provides. The users found the clinician interface simple enough to accomplish simple tasks that simulated the daily applications. The physician user had no criticism and rapidly completed the tasks without assistance. The nurse found some potential flaws in the current design of the interface that could cause potential problems with later widespread application. He found that the data presentation was clear enough. Both users commented on the lack of time-scale harmonization among the displayed parameters, which made the discovery of associations more difficult. One of the users encouraged the development of automated correlation analysis to support the discoveries of potential causative associations. The clinician suggested that the format of motor function data visualization should be in agreement with what they are accustomed to when analyzing the motor data collected by similar systems. The physician would also welcome the incorporation of additional patient-derived data to the EMPARK, such as commenting possibilities for patients to explain their special pattern of symptoms or report on life events (eg, watching an exciting movie that may have caused dyskinesias or delays in medication intake), which could have an impact on either the symptoms or self-management. However, it was also added by both users that such extensions may backfire at the level of patient adherence to EMPARK, as simplicity is often the key for long-term patient cooperation.

Patient Empowerment and the EMPARK system

Using the ICT4PEM [37], we identified the patient empowerment characteristics (PEMs) defined by the model that have clear relevance to and are in accordance with the main goals of EMPARK system development. These PEM characteristics, along with their definitions from ICT4PEM, are listed as follows:

- Control: The ability by which an individual can decide about their level of engagement in the health care process and participate in decisions regarding alternative treatment options, including when these are performed by professionals.
- Understanding: the potential use of the information a patient has regarding their own *health status*, the *diseases*, and the function of the actual and possibly available *health care processes*. Notably, understanding, in our definition, represents the patient's capacity to apply *knowledge* in the specific and individual context of the disease and health care provision. *Information* and its availability to the patient serve as a basis for understanding. Consequently, neither knowledge nor information is sufficient as a characteristic of *empowerment*, although both should be considered as prerequisites for understanding.
- Self-efficacy: the sum of cognitive and physical capabilities possessed by the patient that can be used for self-care.

We were also able to identify several ICT strategies as described by ICT4PEM, which bear obvious relevance to the ICT interventions of EMPARK. In accordance with ICT4PEM, the ICT interventions must target one or several PEM characteristics. Subsequently, we allocated the selected ICT interventions of EMPARK to these intervention categories as described below:

- Feedback
  - Feedback about symptoms, medication, sleep, and activities on a daily and weekly basis
- Monitoring
  - IoT-based home monitoring of disease symptoms, sleep, and self-medication
  - Self-reporting of physical activities and mealtimes
  - Self-reporting of subjective health-related variables
- Communication
  - Providing a visual platform, with disease data used by both clinicians and patients, facilitating mutual communication for co–decision-making.
- Analysis
  - Scoring of symptoms
  - Scoring of physical activities
  - Analysis of self-medication
  - Scoring of sleep
  - Correlation analyses

The sessions with patient organization and questionnaires used to evaluate the design confirmed that these ICT strategies would be welcome for the improvement of the 3 empowerment characteristics we want to improve. However, the degree of effect on empowerment requires evaluation studies that specifically target these characteristics after real-life deployment of EMPARK in clinical studies.

Discussion

Principal Findings

This study presents the latest results from the ongoing system development and evaluation of a novel IoT-based home monitoring system (EMPARK) for people with PD. The EMPARK system represents a unique conceptual approach for both its design and long-term goals. To our knowledge, EMPARK is a unique ICT intervention among people with PD, as it aims to provide patients with feedback about their disease symptoms and disease management. Accordingly, patient empowerment through various means is one of the main goals of the overall research project. The EMPARK system also aims to support the care team through the dynamic visual representation of previously unavailable clinical information and, as a long-term goal, contributes to the academic research.

Primarily, EMPARK is an ICT system that relies largely on IoT tools deployed in the patient's home environment as a source for input data, complemented by patient-related outcome parameters of disease activities and self-management. The objectively measured parameters include continuous daily measures of time spent in bed, motor functions, and medication (levodopa). The self-reported parameters were the time of meals, physical activities, and health-related quality of life questionnaires. As we published information previously, EMPARK synthetizes these parameters with various level of complexity to achieve *daily scores* of sleep quality and duration, motor function, medication compliance, meal timing compliance, and physical activities. Daily scoring of these parameters enables an easy-to-comprehend visualization for patients, along with the potential to provide physicians or nurses with detailed data for deeper analysis. We must emphasize that we have not yet achieved the final goal of implementing a functional and full-scale EMPARK ICT infrastructure in daily practice. However, our current and previous findings provide evidence that the main components of the system, such as databases, computing, and visual interfaces, are almost fully developed and evaluated. Our results from the evaluation of the interfaces highlight directions for potential future developments. For instance, the users suggested automated correlation analysis for discoveries of potential causative associations among the different parameters and to use them either as part of the patient feedback or as clinician support.

Multiple lines of evidence suggest that the EMPARK system will be empowering for people with PD. The ICT design of the EMPARK system followed a user-centered design and patient-centric development. This included iterative steps of development and evaluation with respect to both the patient and clinician interfaces. We recently published a novel framework (ICT4PEM) for ICT interventions, with patient empowerment as the main target [37]. Importantly, the lessons learned through patient-centric design and evaluation of EMPARK made an important contribution to our recognition of the urgent need for such a framework. Retrospective analysis of the ICT design through the principles of ICT4PEM clearly indicates that the system will empower patients. The conceptual elements of patient empowerment can be readily identified as the system

targets. Similarly, ICT interventions within the system fit to those ICT intervention strategies described by the ICT4PEM. Reviewing our previous findings from patient interface evaluations indicates a high level of patient acceptance of the technology and its intuitive use for solving real-life problems common in disease self-management. The content and visual design of the patient interface were well appreciated by the evaluators and in agreement with the previously set requirements during the early developmental stage. Notably, patient feedback from the evaluation led to improvements in the functionalities. We could explain over 80% of the observed interuser variabilities in intention to use by the usual determinants, such as sociodemographic and technology-associated factors. Patient interviews were in accordance with these findings, implicating a high potential for improvement in patient satisfaction, disease control, understanding, and self-efficacy. The latter 3 conceptual elements are also important elements of the patient empowerment paradigm [37]. Taken together, the results of the multistep evaluation with the mock-up patient interface provide us with strong confidence to believe that the selected variables for feedback are neither too little nor too much for improving self-management and ensuring targeted improvements in patient empowerment.

The system also sets the support for the care team as a main project goal. In this study, we describe the user-centric development of a clinician interface and the results of its user-centric evaluation using multiple methodologies. The professional management of patients with PD relies largely on periodic *snapshot* outpatient visits. This renders the care team members to be reliant on patients' memories of their past symptoms in an attempt to obtain a longitudinal picture of the ongoing disease activities and their self-management. Our findings clearly show that care team members, both the physician and nurse, often regard such information as unreliable because of recall bias and find it insufficient for decision-making. Frequent and high-amplitude fluctuations in PD symptoms further emphasize the need for reliable longitudinal follow-up [7,12]. We believe that the application of IoT tools, such as wrist and bed sensors in the patient's home, will be a nonintrusive way to introduce novel data with clinical relevance in the decision-making process [50]. Importantly, the physicians and nurses who participated in the development or the evaluation of the clinician interface share our thoughts. Evaluating care team members think that the parallel availability of information about self-management and disease symptoms will lead to discoveries about the causative associations among them. In addition, our evaluators regarded it crucial for clinical decision-making, as such discoveries could be the basis of a new way of communication and agreement about care plans between the care team members and the patients. In accordance with the suggestions from our clinical interface evaluators, we plan to use the patient interface as the primary tool for shared decision-making for patients and care team members. We obtained evidence regarding the usability and appropriate visual design of the clinician interface. At the same time, we also observed differences that most probably relate to the differential tasks and background knowledge of the different care team members. Clinician interface, in its current form, seems to favor physicians compared with nurses with respect to comprehension and intuitive use. Nevertheless, the low number of evaluators should be regarded as a limitation as such differences in user experience may also be explained by other factors. We are currently undertaking efforts to expand the number of participants in the clinician interface evaluation to better understand the determining factors of user acceptance.

Research support is also one of the long-term goals of the EMPARK system. We now present clear support for such potential as part of the evaluation of the clinician interface. The evaluators of the clinician interface support our view that the EMPARK infrastructure, with a special focus on the newly defined scores of daily PD-related activities, could serve the purposes of observational or randomized controlled trials in the future. Our findings strongly support the incorporation of the scored parameters developed by the system into clinical trials, which can be justified by their relevance to the clinical decisions, lack of availability via other sources, and the patients' own judgment about their usefulness. Such trials could help to establish the role of a parameter scoring system in the daily care of patients with PD, either after or partially in parallel with the previously described stages of implementation. In addition, it would also allow the scientific evaluation of a novel care model for patients with PD, with a focus shifted from snapshot controls to longitudinal assessment and mutual care plan implementation.

Limitation and Future Work

There are several limitations with respect to the generalizability of our results, which are not only with regard to the potential user acceptance of a clinician interface in a large-scale trial.

We are aware that the implementation of the EMPARK system is crucial for assessing long-term user acceptance. Evaluators of the clinician interface and professionals with extensive experience in PD care consider the progressive neurological deficits of patients with PD as a potential risk for a diminishing system implication in the home environment over a longer period. Although we agree, we believe this risk is low considering our results that indicate a high level of acceptance and assessed usefulness among patient evaluators. Similarly, a conclusive assessment of patient empowerment and enhanced self-management, which are our main goals, awaits real-life implementation of the EMPARK system.

To address these issues and evaluate strategic goals, we are currently undertaking steps toward the experimental implementation of the EMPARK system in a real-life milieu. Early, limited implementation (ie, single-institute evaluation) can follow the design research principles with the potential of an ongoing, on-demand, smaller adjustment of the system architecture. However, the main assessment of long-term goals, such as the effect of the EMPARK system on patient empowerment, overall health status, or auxiliary effects on health care use, would require overall system stability during a larger clinical trial. We suggest that such a clinical trial should also address the role of the scored parameters as tools for daily clinical practice in PD care. Implementation at a larger scale would result in the accumulation of a large amount of clinical data. We believe that the great potential of such a database for auxiliary research can facilitate the implementation of the EMPARK system by health care providers.

Conclusions

The EMPARK system is a novel IoT-based home monitoring system for providing patients with PD and team members with feedback about disease activities. EMPARK is unique in its primary aim of providing patients with feedback on disease symptoms and self-management, consequently enhancing patient empowerment. We describe the user-centric development and evaluation of the system's clinical interface and show results indicating its high usability in clinical management and shared decision-making with patients. We also provide extensive evidence of how the EMPARK system infrastructure would empower patients and suggest future application in PD-related research. Our results represent the last phase in development before the implementation of EMPARK in a real-life environment.

Acknowledgments

This research was supported by the Swedish Knowledge Foundation (grant 20160176) and Sensidose AB, Cenvigo AB, and Nethouse Sverige AB. The authors would like to thank the Newbreed EU Cofund doctoral program within the focus area of Successful Ageing at Örebro University for funding LK.

Conflicts of Interest

None declared.

Multimedia Appendix 1
First task-based evaluation.
[DOCX File , 14 KB - formative_v6i6e31485_app1.docx]

Multimedia Appendix 2
Second task-based evaluation.
[DOCX File , 15 KB - formative_v6i6e31485_app2.docx]

Multimedia Appendix 3
Semistructured interview questions.
[DOCX File , 14 KB - formative_v6i6e31485_app3.docx]

References

1. Bloem B, Stocchi F. Move for change part I: a European survey evaluating the impact of the EPDA Charter for People with Parkinson's disease. Eur J Neurol 2012 Mar 04;19(3):402-410 [FREE Full text] [doi: 10.1111/j.1468-1331.2011.03532.x] [Medline: 21967281]
2. Muzerengi S, Contrafatto D, Chaudhuri KR. Non-motor symptoms: identification and management. Parkinsonism Related Disorders 2007;13:S450-S456. [doi: 10.1016/s1353-8020(08)70048-8]
3. Visser M, van Rooden SM, Verbaan D, Marinus J, Stiggelbout AM, van Hilten JJ. A comprehensive model of health-related quality of life in Parkinson's disease. J Neurol 2008 Oct 24;255(10):1580-1587. [doi: 10.1007/s00415-008-0994-4] [Medline: 18821041]
4. Stamford JA, Schmidt PN, Friedl KE. What Engineering Technology Could Do for Quality of Life in Parkinson's Disease: A Review of Current Needs and Opportunities. IEEE J Biomed Health Inform 2015 Nov;19(6):1862-1872. [doi: 10.1109/JBHI.2015.2464354] [Medline: 26259205]
5. Dorsey ER, Vlaanderen FP, Engelen LJ, Kieburtz K, Zhu W, Biglan KM, et al. Moving Parkinson care to the home. Mov Disord 2016 Sep 08;31(9):1258-1262 [FREE Full text] [doi: 10.1002/mds.26744] [Medline: 27501323]
6. Bekris LM, Mata IF, Zabetian CP. The genetics of Parkinson disease. J Geriatr Psychiatry Neurol 2010 Dec 11;23(4):228-242 [FREE Full text] [doi: 10.1177/0891988710383572] [Medline: 20938043]
7. Holford N, Nutt JG. Disease progression, drug action and Parkinson's disease: why time cannot be ignored. Eur J Clin Pharmacol 2008 Feb 19;64(2):207-216 [FREE Full text] [doi: 10.1007/s00228-007-0427-9] [Medline: 18092155]
8. Stocchi F, Bloem BR. Move for Change Part II: a European survey evaluating the impact of the EPDA Charter for people with Parkinson's disease. Eur J Neurol 2013 Mar 03;20(3):461-472 [FREE Full text] [doi: 10.1111/j.1468-1331.2012.03876.x] [Medline: 23034057]
9. Papapetropoulos S, Mitsi G, Espay AJ. Digital health revolution: is it time for affordable remote monitoring for Parkinson's disease? Front Neurol 2015 Feb 26;6:34 [FREE Full text] [doi: 10.3389/fneur.2015.00034] [Medline: 25767462]
10. Maetzler W, Domingos J, Srulijes K, Ferreira JJ, Bloem BR. Quantitative wearable sensors for objective assessment of Parkinson's disease. Mov Disord 2013 Oct 12;28(12):1628-1637. [doi: 10.1002/mds.25628] [Medline: 24030855]
11. Wicks P, Stamford J, Grootenhuis MA, Haverman L, Ahmed S. Innovations in e-health. Qual Life Res 2014 Feb 14;23(1):195-203 [FREE Full text] [doi: 10.1007/s11136-013-0458-x] [Medline: 23852096]

12. Espay AJ, Bonato P, Nahab FB, Maetzler W, Dean JM, Klucken J, Movement Disorders Society Task Force on Technology. Technology in Parkinson's disease: callenges and opportunities. Mov Disord 2016 Sep 29;31(9):1272-1282 [FREE Full text] [doi: 10.1002/mds.26642] [Medline: 27125836]
13. Rovini E, Esposito D, Maremmani C, Bongioanni P, Cavallo F. Empowering patients in self-management of Parkinson's disease through cooperative ICT systems. In: Wearable Technologies: Concepts, Methodologies, Tools, and Applications. Pennsylvania, United States: IGI Global; 2018:637-663.
14. Bhidayasiri R, Mari Z. Digital phenotyping in Parkinson's disease: empowering neurologists for measurement-based care. Parkinsonism Relat Disord 2020 Nov;80:35-40. [doi: 10.1016/j.parkreldis.2020.08.038] [Medline: 32937225]
15. Hansen C, Sanchez-Ferro A, Maetzler W. How Mobile Health Technology and Electronic Health Records Will Change Care of Patients with Parkinson's Disease. J Parkinsons Dis 2018;8(s1):S41-S45 [FREE Full text] [doi: 10.3233/JPD-181498] [Medline: 30584169]
16. Pasluosta CF, Gassner H, Winkler J, Klucken J, Eskofier BM. An Emerging Era in the Management of Parkinson's Disease: Wearable Technologies and the Internet of Things. IEEE J Biomed Health Inform 2015 Nov;19(6):1873-1881. [doi: 10.1109/JBHI.2015.2461555] [Medline: 26241979]
17. Klucken J, Friedl KE, Eskofier BM, Hausdorff JM. Guest Editorial: Enabling Technologies for Parkinson's Disease Management. IEEE J Biomed Health Inform 2015 Nov;19(6):1775-1776. [Medline: 26866144]
18. Klucken J, Krüger R, Schmidt P, Bloem BR. Management of Parkinson's Disease 20 Years from Now: Towards Digital Health Pathways. J Parkinsons Dis 2018;8(s1):S85-S94 [FREE Full text] [doi: 10.3233/JPD-181519] [Medline: 30584171]
19. Marsili L, Bologna M, Miyasaki JM, Colosimo C. Parkinson's disease advanced therapies - A systematic review: more unanswered questions than guidance. Parkinsonism Relat Disord 2021 Feb;83:132-139. [doi: 10.1016/j.parkreldis.2020.10.042] [Medline: 33158747]
20. Walls JG, Widmeyer GR, El Sawy OA. Building an information system design theory for vigilant EIS. Inf Syst Res 1992 Mar;3(1):36-59. [doi: 10.1287/isre.3.1.36]
21. Senek M, Hellström M, Albo J, Svenningsson P, Nyholm D. First clinical experience with levodopa/carbidopa microtablets in Parkinson's disease. Acta Neurol Scand 2017 Dec 15;136(6):727-731. [doi: 10.1111/ane.12756] [Medline: 28299787]
22. Wang H, Jafari R, Zhou G, Venkatasubramanian KK, Stella Sun J, Bonato P, et al. Guest editorial : special issue on internet of things for smart and connected health. IEEE Internet Things J 2015 Feb;2(1):1-4. [doi: 10.1109/jiot.2015.2394696]
23. Ozinga SJ, Machado AG, Miller Koop M, Rosenfeldt AB, Alberts JL. Objective assessment of postural stability in Parkinson's disease using mobile technology. Mov Disord 2015 Aug 25;30(9):1214-1221. [doi: 10.1002/mds.26214] [Medline: 25809137]
24. Horak FB, Mancini M. Objective biomarkers of balance and gait for Parkinson's disease using body-worn sensors. Mov Disord 2013 Sep 15;28(11):1544-1551 [FREE Full text] [doi: 10.1002/mds.25684] [Medline: 24132842]
25. Mariani B, Jiménez MC, Vingerhoets FJ, Aminian K. On-shoe wearable sensors for gait and turning assessment of patients with Parkinson's disease. IEEE Trans Biomed Eng 2013 Jan;60(1):155-158. [doi: 10.1109/TBME.2012.2227317] [Medline: 23268531]
26. Cavanaugh J, Ellis T, Earhart G, Ford M, Foreman K, Dibble L. Capturing ambulatory activity decline in Parkinson's disease. J Neurol Phys Ther 2012 Jun;36(2):51-57 [FREE Full text] [doi: 10.1097/NPT.0b013e318254ba7a] [Medline: 22592060]
27. Rahimi F, Bee C, Duval C, Boissy P, Edwards R, Jog M, EMAP Group. Using ecological whole body kinematics to evaluate effects of medication adjustment in Parkinson disease. J Parkinsons Dis 2014;4(4):617-627. [doi: 10.3233/JPD-140370] [Medline: 25055960]
28. Schlachetzki JC, Barth J, Marxreiter F, Gossler J, Kohl Z, Reinfelder S, et al. Wearable sensors objectively measure gait parameters in Parkinson's disease. PLoS One 2017 Oct 11;12(10):e0183989 [FREE Full text] [doi: 10.1371/journal.pone.0183989] [Medline: 29020012]
29. Marxreiter F, Gaßner H, Borozdina O, Barth J, Kohl Z, Schlachetzki JC, et al. Sensor-based gait analysis of individualized improvement during apomorphine titration in Parkinson's disease. J Neurol 2018 Nov 8;265(11):2656-2665. [doi: 10.1007/s00415-018-9012-7] [Medline: 30196324]
30. Mera TO, Heldman DA, Espay AJ, Payne M, Giuffrida JP. Feasibility of home-based automated Parkinson's disease motor assessment. J Neurosci Methods 2012 Jan 15;203(1):152-156 [FREE Full text] [doi: 10.1016/j.jneumeth.2011.09.019] [Medline: 21978487]
31. Rovini E, Maremmani C, Cavallo F. Automated systems based on wearable sensors for the management of Parkinson's disease at home: a systematic review. Telemed J E Health 2019 Mar;25(3):167-183. [doi: 10.1089/tmj.2018.0035] [Medline: 29969384]
32. Memedi M, Tshering G, Fogelberg M, Jusufi I, Kolkowska E, Klein G. An interface for IoT: feeding back health-related data to Parkinson's disease patients. J Sense Actuator Network 2018 Mar 12;7(1):14. [doi: 10.3390/jsan7010014]
33. Memedi M, Lindqvist J, Tunedal T, Duvåker A. A study on pre-adoption of a self-management application by Parkinson's disease patients. In: Proceedings of the 2018 International Congress. 2018 Presented at: 2018 International Congress; Oct 5-9, 2018; Hong Kong.
34. Nieuwboer A, De Weerdt W, Dom R, Lesaffre E. A frequency and correlation analysis of motor deficits in Parkinson patients. Disabil Rehabil 1998 Apr 28;20(4):142-150. [doi: 10.3109/09638289809166074] [Medline: 9571381]

35. Lakshminarayana R, Wang D, Burn D, Chaudhuri KR, Cummins G, Galtrey C, SMART-PD Investigators. Smartphone- and internet-assisted self-management and adherence tools to manage Parkinson's disease (SMART-PD): study protocol for a randomised controlled trial (v7; 15 August 2014). Trials 2014 Sep 25;15(1):374 [FREE Full text] [doi: 10.1186/1745-6215-15-374] [Medline: 25257518]
36. Fleisher JE, Shah K, Fitts W, Dahodwala NA. Associations and implications of low health literacy in Parkinson's disease. Mov Disord Clin Pract 2016 Dec 14;3(3):250-256 [FREE Full text] [doi: 10.1002/mdc3.12272] [Medline: 27331078]
37. Karni L, Dalal K, Memedi M, Kalra D, Klein GO. Information and communications technology-based interventions targeting patient empowerment: framework development. J Med Internet Res 2020 Aug 26;22(8):e17459 [FREE Full text] [doi: 10.2196/17459] [Medline: 32845245]
38. Peffers K, Tuunanen T, Rothenberger MA, Chatterjee S. A design science research methodology for information systems research. J Manag Inf Syst 2014 Dec 08;24(3):45-77. [doi: 10.2753/mis0742-1222240302]
39. Abras C, Maloney-Krichmar D, Preece J. User-centered design. In: Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications; 2004.
40. Espay AJ, Hausdorff JM, Sánchez-Ferro Á, Klucken J, Merola A, Bonato P, Movement Disorder Society Task Force on Technology. A roadmap for implementation of patient-centered digital outcome measures in Parkinson's disease obtained using mobile health technologies. Mov Disord 2019 May 22;34(5):657-663 [FREE Full text] [doi: 10.1002/mds.27671] [Medline: 30901495]
41. Williams JA, Cook DJ. Forecasting behavior in smart homes based on sleep and wake patterns. Technol Health Care 2017 Feb 21;25(1):89-110. [doi: 10.3233/thc-161255]
42. Jenkinson C, Fitzpatrick R, Peto V, Greenhall R, Hyman N. The PDQ-8: Development and validation of a short-form parkinson's disease questionnaire. Psychology & Health 1997 Dec;12(6):805-814. [doi: 10.1080/08870449708406741]
43. van Hout B, Janssen M, Feng Y, Kohlmann T, Busschbach J, Golicki D, et al. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. Value Health 2012 Jul;15(5):708-715 [FREE Full text] [doi: 10.1016/j.jval.2012.02.008] [Medline: 22867780]
44. Wilson C. Brainstorming and Beyond A User-Centered Design Method. Oxford: Morgan Kaufman; 2013.
45. Jaspers MW. A comparison of usability methods for testing interactive health technologies: methodological aspects and empirical evidence. Int J Med Inform 2009 May;78(5):340-353. [doi: 10.1016/j.ijmedinf.2008.10.002] [Medline: 19046928]
46. Nielsen J. How to conduct a heuristic evaluation. Nielsen Norman Group. URL: https://www.nngroup.com/articles/how-to-conduct-a-heuristic-evaluation/ [accessed 2022-05-09]
47. Lewis JR. IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. Int J Human Comput Interact 1995 Jan;7(1):57-78. [doi: 10.1080/10447319509526110]
48. Visualization of Time-Oriented Data. London: Springer; 2011.
49. Farzanehfar P, Horne M. Evaluation of the Parkinson's KinetiGraph in monitoring and managing Parkinson's disease. Expert Rev Med Devices 2017 Aug 06;14(8):583-591. [doi: 10.1080/17434440.2017.1349608] [Medline: 28675078]
50. Almogren A. An automated and intelligent Parkinson disease monitoring system using wearable computing and cloud technology. Cluster Comput 2018 Jan 5;22(S1):2309-2316. [doi: 10.1007/s10586-017-1591-z]

Abbreviations

EQ-5D-3L: European Quality of Life 5 Dimensions 3 Level
ICT: information communication technology
ICT4PEM: ICT for patient empowerment model
IoT: Internet of Things
PD: Parkinson disease
PEM: patient empowerment model

Edited by A Mavragani; submitted 23.06.21; peer-reviewed by S Mantalvanos, E Toki, L Heryawan; comments to author 24.02.22; revised version received 08.03.22; accepted 18.03.22; published 09.06.22.

Please cite as:
Karni L, Jusufi I, Nyholm D, Klein GO, Memedi M
Toward Improved Treatment and Empowerment of Individuals With Parkinson Disease: Design and Evaluation of an Internet of Things System
JMIR Form Res 2022;6(6):e31485
URL: https://formative.jmir.org/2022/6/e31485
doi:10.2196/31485
PMID:35679097

Short Paper

Acceptability and Feasibility of Peer-to-Peer Text Messaging Among Adolescents to Increase Clinic Visits and Sexually Transmitted Infection Testing: Interrupted Times-Series Analysis

Marguerita Lightfoot[1], PhD; Joi Jackson-Morgan[2], MPH; Lance Pollack[1], PhD; Ayanna Bennett[2], MD

[1]Division of Prevention Science, Department of Medicine, University of California San Francisco, San Francisco, CA, United States
[2]3rd Street Youth Center and Clinic, San Francisco, CA, United States

Corresponding Author:
Marguerita Lightfoot, PhD
Division of Prevention Science
Department of Medicine
University of California San Francisco
550 16th Street, 3rd Floor
San Francisco, CA, 94143
United States
Phone: 1 415 502 4320
Email: lightfom@ohsu.edu

Abstract

Background: Adolescents are disproportionately affected by sexually transmitted infections (STIs), including HIV. Many youths with asymptomatic STI or related symptoms do not seek treatment and may not be screened if accessing the health care system for other reasons.

Objective: We examined intervention completion and changes in the number of new patients, the number of STI or HIV tests, and the sexual risk profile of patients over time to determine the feasibility and acceptability of a peer-driven text messaging strategy to connect youth to STI and HIV services.

Methods: The intervention enlisted consecutive patients at an adolescent medicine clinic to send a text message to 5 peers they believed were sexually active and lived in the clinic's service area. The intervention was evaluated using an interrupted time-series design in which baseline clinic service levels were documented during a 35-week lead-in period, followed by a 20-week intervention implementation period, and a 16-week period of continued clinic observation. Clinic and patient data were obtained through chart abstraction from intake forms that occurred during the entire study period. Analyses conducted in 2015 used a generalized linear mixed model.

Results: Of the 153 patients approached to participate, 100 agreed to send SMS text messages. Most (n=55, 55%) reported no concerns with sending the text message. No adverse events or negative outcomes were reported. Adolescent STI testing, positive test results, and reported risk behavior increased post intervention, although this was not statistically significant, likely because of the small sample size.

Conclusions: Given low youth uptake of health care services, and STI/HIV screening, in particular, new strategies are needed to address access barriers. Common approaches for reaching youth are resource-intensive and often miss those not connected to school or community programs. The peer-based text messaging strategy showed promise for both increasing the number of youths accessing health services and finding youths engaging in sexual risk behaviors and most in need of sexual health screening and services.

(JMIR Form Res 2022;6(6):e32416) doi:10.2196/32416

KEYWORDS

HIV prevention; STI prevention; adolescents; youth; text messaging; SMS; peer-to-peer intervention; HIV; STI; HIV testing

Introduction

Adolescents are disproportionately affected by sexually transmitted infections (STIs). The World Health Organization estimates that 50% of the 30 million HIV infections worldwide occurred in youths between the ages of 15 and 24 years [1]. In the United States, almost one-quarter of new HIV infections are among youth and only 22% of sexually active youth have tested for HIV [2]. Furthermore, approximately half of all cases of STIs occur in people between the ages of 15 and 24 years [3]. However, many youths have asymptomatic STIs or do not seek services when they do have symptoms [4-6]. Youths who do access the health care system may not be screened for HIV [4,7]. Thus, despite the need for STI and HIV services, adolescents are among the most medically underserved.

Adolescents face barriers to accessing sexual health services, including stigma, low knowledge about STIs and available services, cost, clinic wait times, scheduling conflicts, and embarrassment [8-11]. Further, adolescents are least likely to have access to health care in the United States [12-14]. Hence, there has been a call for innovative solutions to increase adolescent STI/HIV screening [15]. In overcoming barriers to STI/HIV testing and access to health care, a peer-driven model may be particularly effective. Peers have the unique advantage of speaking from their own experiences; they also have credibility in delivering messages, as well as linguistic and cultural familiarity [16] and access to those in their social and sexual networks who are disconnected from the health care system. Peer education and outreach strategies have been successful in increasing the use of health resources in adolescents [17]. Further, SMS text messaging has proven successful in engaging STI/HIV screening among youth [18-20]. A recent systematic review found that text message reminders were most successful in increasing rates of HIV testing compared to other strategies [21]. We designed a peer-based approach that used text messaging to increase STI/HIV screening among adolescents. We report the feasibility and acceptability of this strategy.

Methods

Study Design

We used an interrupted time-series design to examine the impact of the text message intervention, deployed in an adolescent medicine clinic serving youth aged 12 to 24 years. Baseline service levels were documented during a 35-week lead-in period (weeks 1-35). This was followed by a 20-week period when patients were asked to send text messages to 5 friends who they believed were sexually active and lived in the clinic's service area (weeks 36-55). Following the intervention was a 16-week period of continued clinic observation (weeks 56-72). Chart abstraction from intake forms continued during the entire study (weeks 1-72).

Ethics Approval

The study was approved by the University of California, San Francisco, Institutional Review Board (approval #12-08516).

Data Collection

Our target was 100 participants to ascertain feasibility and acceptability. Of the 153 patients approached by research staff, 28 (18%) refused, 25 (16%) did not have a mobile phone with them, and 100 (65%) provided informed consent and sent text messages. Participants were told that the goal of the text messages was to encourage their friends to visit the clinic and receive STI screening. Participants were provided a text messaging guide that reviewed considerations for developing messages (eg, let your friend know you care about them). Participants developed their own message. Subsequent to sending the messages, participants were provided US $10 for completing a short posttexting interview. The text message and responses from friends while the participant was still in the clinic were recorded by research staff. Staff recontacted participants 1 week after their clinic visit to ascertain issues with friends following the intervention (n=93).

Data Analysis

Analyses compared the 35-week "pretexting" data to the 36-week postintervention period. Additional comparisons accounted for seasonal fluctuations in service levels by comparison to the same dates in the preceding year. We examined changes in the number of new patients seen (including patients not seen for at least 12 months), the number of STI tests conducted and positive results, the number of HIV tests conducted and positive results, and the number of at-risk patients seen at the clinic. Patients were characterized as at risk if they reported multiple partners, infrequent condom use, or had an STI at intake. Counts for each outcome were aggregated into average monthly totals, and analyses were performed using a generalized linear mixed model. Outcomes were modeled as overdispersed Poisson distribution variables. For clarity, only the mean of the counts during each time period was provided.

Results

Peer Texting

All participants sent the required 5 messages (see Textbox 1 for examples of messages sent and received), and 18% (n=18) of participants sent more than 5 messages. In the brief posttexting interview, most patients (n=55, 55%) reported no concerns with sending the text message, 18% (n=18) expressed worry about their friends' response, and 27% (n=27) expressed concern that their friend would not heed their advice (eg, their friend would not call the clinic or would not listen). In the 1-week postmessaging interview, no youth reported adverse events or negative outcomes from sending the message.

Textbox 1. Examples of text messages sent and received by participants.

Example messages sent by participants

- "Hey girl this _____ wen u cum bck frm Texas here at the clinic on [redacted] u shuld stp bi they hav gud classes n program n ull fel lyk home"
- "Hey. So I'm at the clinic and I'm gonna get tested. You should come get checked too. I'll come with you if you want. It's over on [redacted] next to the new WingStop."
- "Hey I'm at the [redacted] it's a good idea to get checked out"
- "I just got checked out at the [redacted] and they were really nice they offer free testing and helped me a lot, you should come down too."

Example messages received from peers

- "I just got check, that shit ain't nothing to be played with"
- "Lmfao. Nice. Okay well when I get home"
- "Fa sho. What time it ends? I get out @ 4"
- "Bitch what you telling me for? I go to my dr. every 3 months for the free and I damn sho aint got HIV"

Clinic Attendance and STI/HIV Testing

A total of 430 new patients were seen at the clinic during the 72 weeks of data collection, of whom 338 (79%) were female. New patients were 61% (n=262) African American and 17% (n=73) Hispanic, had a mean age of 19.6 (SD 2.6, range 13-25) years, and primarily identified as heterosexual (n=375, 87%). Almost half of the patients reported they had no regular doctor (n=185, 43%) or health care (n=194, 45%).

In a brief posttexting interview with the 100 participants, most patients (n=55, 55%) reported no concerns with sending the text message, 18% (n=18) expressed worry about their friends' response, and 27% (n=27) expressed a concern that their friend would not heed their advice. In a follow-up interview 1 week after messaging, no youth reported adverse events or negative outcomes from sending the message.

We examined changes in the number of new patients seen, the number of STI tests conducted and positive results, the number of HIV tests conducted and positive results, and the number of at-risk patients seen at the clinic. Patients were characterized as at risk if they reported multiple partners, infrequent condom use, or any STI infection. All clinic services outcomes increased during the posttexting period compared to the pretexting period (Table 1), including weekly STI tests (4.11 vs 5.18, *P*=.58), weekly number of youths testing positive for STI (0.60 vs 0.76, *P*=.17), weekly number of youths tested for HIV (0.91 vs 1.35, *P*=.48), and weekly number of patients reporting high-risk sexual behaviors (4.89 vs 5.29, *P*=.55). However, these findings did not achieve statistical significance. In order to account for possible seasonal and period effects, we also examined the means for a subset of the preintervention phase that matched the calendar months covered in the postintervention phase but occurred 1 year earlier. Findings were similar, suggesting that the variation in average patient counts between the pre- and postintervention periods was not due to seasonal variation.

Table 1. Average weekly patient count during the preintervention, intervention, and postintervention phases.

Variable	Weekly patient count[a], mean (SD)			
	Preintervention (weeks 1-35)	Calendar-matched preintervention (weeks 6-23)	Intervention (weeks 36-55)	Postintervention (weeks 56-72)
Total	6.00 (3.25)	5.50 (2.83)	5.40 (2.11)	6.59 (2.79)
Tested for STI[b]	4.11 (2.86)	3.50 (2.50)	3.95 (2.16)	5.18 (2.13)
Positive for STI	0.60 (.081)	0.39 (0.61)	0.60 (0.88)	0.76 (0.90)
Tested for HIV	0.91 (1.72)	0.50 (0.71)	0.85 (0.88)	1.35 (1.69)
High sexual risk[c]	4.89 (2.80)	4.28 (2.32)	4.15 (1.95)	5.29 (2.34)

[a]Count of new patients and re-engaged patients (patients not seen for at least 12 months).

[b]STI: sexually transmitted infection.

[c]Self-reported as having ≥2 sex partners in the prior 3 months, or self-reported an STI diagnosis in the past year, or tested positive for gonorrhea or chlamydia at the clinic visit, or self-reported not always using a condom when having sexual intercourse.

Discussion

Key Findings

Given the low number of youths accessing health care services and STI/HIV screening, new strategies are needed to address the barriers that exist. We found that text messaging between peers is a feasible and acceptable strategy with potential for increasing STI/HIV testing. We did not provide standard messages; rather, we were interested in whether youth were able to develop their own persuasive messages. Most youths agreed to send a message and did not express concern or experience complaints from peers. One in 5 participants sent more than the required number of messages. The approach appears feasible and acceptable. The simplicity and low-resource requirements of this approach also make it more sustainable and significantly less costly than recent efforts to build smartphone apps, social media campaigns, or other technology-based interventions to increase testing [18].

We were unable to examine if messages were read or to directly ask recipients about their reactions. Given that over 97% of text messages are opened and 90% are read within 3 minutes of their delivery [22], we did record the immediate response from message recipients. There were few negative text responses, and no negative consequences or adverse events reported by participants during the follow-up interviews. This suggests that the intervention may also be acceptable to the recipients of the message.

Limitations

This study was a proof of concept, and, thus, has limitations in interpretation. The study had a small sample size and was conducted at a single adolescent medical clinic. There may have been selection bias where youths who were more comfortable discussing sexual health with their friends may be overrepresented. However, it is likely that a formal implementation of this intervention would also be on an opt-in basis, strengthening the external validity of our pilot.

Conclusions

While this pilot study was not powered to detect significant differences over time, we did find encouraging increases in the number of weekly STI tests conducted, the number of youths testing positive monthly for STIs, the number of youths tested monthly for HIV, and the number of patients reporting sexual risk behaviors. Thus, text messaging may have increased the number of youths accessing services and STI/HIV screening and reached those engaging in risky sexual behavior who are most in need of screening. This is particularly noteworthy given the clinic primarily serves a low-income population of Black and Hispanic adolescents, a population disproportionately impacted by STI and HIV [23].

Conflicts of Interest

None declared.

References

1. Adolescents: Health risks and solutions. World Health Organization. 2018 Jan 18. URL: https://www.who.int/news-room/fact-sheets/detail/adolescents-health-risks-and-solutions [accessed 2022-05-16]
2. HIV in the United States by age. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/hiv/group/age/youth/index.html [accessed 2022-05-16]
3. Sexually transmitted diseases: Adolescents and young adults. Centers for Disease Control and Prevention. 2018. URL: https://www.cdc.gov/std/life-stages-populations/adolescents-youngadults.htm [accessed 2022-05-16]
4. Ellen J, Lane M, McCright J. Are adolescents being screened for sexually transmitted diseases? A study of low income African American adolescents in San Francisco. Sex Transm Infect 2000 Apr;76(2):94-97 [FREE Full text] [doi: 10.1136/sti.76.2.94] [Medline: 10858709]
5. Fortenberry JD. Health care seeking behaviors related to sexually transmitted diseases among adolescents. Am J Public Health 1997 Mar;87(3):417-420. [doi: 10.2105/ajph.87.3.417] [Medline: 9096544]
6. Marcell AV, Morgan AR, Sanders R, Lunardi N, Pilgrim NA, Jennings JM, et al. The socioecology of sexual and reproductive health care use among young urban minority males. J Adolesc Health 2017 Apr;60(4):402-410 [FREE Full text] [doi: 10.1016/j.jadohealth.2016.11.014] [Medline: 28065520]
7. Adebayo OW, Gonzalez-Guarda RM. Factors associated with HIV testing in youth in the United States: an integrative review. J Assoc Nurses AIDS Care 2017 May;28(3):342-362. [doi: 10.1016/j.jana.2016.11.006] [Medline: 27993497]
8. Bender SS, Fulbright YK. Content analysis: a review of perceived barriers to sexual and reproductive health services by young people. Eur J Contracept Reprod Health Care 2013 Jun 26;18(3):159-167. [doi: 10.3109/13625187.2013.776672] [Medline: 23527736]
9. Garney W, Wilson K, Ajayi KV, Panjwani S, Love SM, Flores S, et al. Social-ecological barriers to access to healthcare for adolescents: a scoping review. Int J Environ Res Public Health 2021 Apr 14;18(8):4138 [FREE Full text] [doi: 10.3390/ijerph18084138] [Medline: 33919813]
10. Hock-Long L, Herceg-Baron R, Cassidy AM, Whittaker PG. Access to adolescent reproductive health services: financial and structural barriers to care. Perspect Sexual Reprod Health 2003 May;35(3):144-147. [doi: 10.1111/j.1931-2393.2003.tb00135.x]

11. Tilson EC, Sanchez V, Ford CL, Smurzynski M, Leone PA, Fox KK, et al. Barriers to asymptomatic screening and other STD services for adolescents and young adults: focus group discussions. BMC Public Health 2004 Jun 09;4(1):21 [FREE Full text] [doi: 10.1186/1471-2458-4-21] [Medline: 15189565]
12. Another look at health insurance coverage rates for young adults. The United States Census Bureau. 2018 Sep 16. URL: https://www.census.gov/newsroom/blogs/random-samplings/2018/09/another_look_at_heal.html [accessed 2022-05-16]
13. Spencer DL, McManus M, Call KT, Turner J, Harwood C, White P, et al. Health care coverage and access among children, adolescents, and young adults, 2010-2016: implications for future health reforms. J Adolesc Health 2018 Jun;62(6):667-673 [FREE Full text] [doi: 10.1016/j.jadohealth.2017.12.012] [Medline: 29599046]
14. Adolescent health—Volume I: Summary and policy options, OTA-H-468. US Congress, Office of Technology Assessment. Washington, DC: US Government Printing Office; 1991 Apr. URL: https://ota.fas.org/reports/9102.pdf [accessed 2022-05-16]
15. Fernandez MI, Harper GW, Hightow-Weidman LB, Kapogiannis BG, Mayer KH, Parsons JT, et al. Research priorities to end the adolescent HIV epidemic in the United States: viewpoint. JMIR Res Protoc 2021 Jan 04;10(1):e22279 [FREE Full text] [doi: 10.2196/22279] [Medline: 33393918]
16. Miller-Day M, Hecht ML. Narrative means to preventative ends: a narrative engagement framework for designing prevention interventions. Health Commun 2013 Oct;28(7):657-670 [FREE Full text] [doi: 10.1080/10410236.2012.762861] [Medline: 23980613]
17. Boyer CB, Sieverding J, Siller J, Gallaread A, Chang YJ. Youth United Through Health Education: community-level, peer-led outreach to increase awareness and improve noninvasive sexually transmitted infection screening in urban African American youth. J Adolesc Health 2007 Jun;40(6):499-505. [doi: 10.1016/j.jadohealth.2006.09.020] [Medline: 17531755]
18. Miller RL, Boyer CB, Chiaramonte D, Lindeman P, Chutuape K, Cooper-Walker B, et al. Evaluating testing strategies for identifying youths with HIV infection and linking youths to biomedical and other prevention services. JAMA Pediatr 2017 Jun 01;171(6):532-537 [FREE Full text] [doi: 10.1001/jamapediatrics.2017.0105] [Medline: 28418524]
19. Saragih ID, Imanuel Tonapa S, Porta CM, Lee B. Effects of telehealth interventions for adolescent sexual health: A systematic review and meta-analysis of randomized controlled studies. J Telemed Telecare 2021 Dec 13:1357633X2110477. [doi: 10.1177/1357633x211047762]
20. Miller RL, Chiaramonte D, Strzyzykowski T, Sharma D, Anderson-Carpenter K, Fortenberry JD. Improving timely linkage to care among newly diagnosed HIV-infected youth: results of SMILE. J Urban Health 2019 Dec 01;96(6):845-855 [FREE Full text] [doi: 10.1007/s11524-019-00391-z] [Medline: 31677014]
21. Paschen-Wolff MM, Restar A, Gandhi AD, Serafino S, Sandfort T. A systematic review of interventions that promote frequent HIV testing. AIDS Behav 2019 Apr 1;23(4):860-874 [FREE Full text] [doi: 10.1007/s10461-019-02414-x] [Medline: 30707329]
22. Conversational advertising. Mobilesquared / SinglePoint. 2010 Jun. URL: https://mobilesquared.co.uk/portfolio/conversational-advertising/ [accessed 2022-05-15]
23. Sexually transmitted disease surveillance 2018. Centers for Disease Control and Prevention. 2019. URL: https://www.cdc.gov/std/stats18/STDSurveillance2018-full-report.pdf [accessed 2022-05-13]

Abbreviations

STI: sexually transmitted infection

Edited by A Mavragani; submitted 27.07.21; peer-reviewed by A Sharma, P Serrano; comments to author 22.01.22; revised version received 05.04.22; accepted 11.05.22; published 09.06.22.

Please cite as:
Lightfoot M, Jackson-Morgan J, Pollack L, Bennett A
Acceptability and Feasibility of Peer-to-Peer Text Messaging Among Adolescents to Increase Clinic Visits and Sexually Transmitted Infection Testing: Interrupted Times-Series Analysis
JMIR Form Res 2022;6(6):e32416
URL: https://formative.jmir.org/2022/6/e32416
doi:10.2196/32416
PMID:35686737

Original Paper

General Demographics and Behavioral Patterns of Visitors Using a Self-help Website for Identification of and Intervention in Alcoholism and Common Mental Disorders in Suriname: Descriptive Study

Raj Jadnanansing[1,2], MSc; Jack Dekker[3,4], Prof Dr; Kajal Etwaroo[1,2], BSc; Rudi Dwarkasing[1,5], MD; Vincent Lumsden[1,5], MD; Robbert Bipat[5*], Prof Dr; Matthijs Blankers[4*], PhD

[1]Psychiatrisch Centrum Suriname, Paramaribo, Suriname
[2]Faculty of Social Science, Anton de Kom University of Suriname, Paramaribo, Suriname
[3]Vrije Universiteit Amsterdam, Amsterdam, Netherlands
[4]Arkin, Amsterdam, Netherlands
[5]Faculty of Medical Science, Anton de Kom University of Suriname, Paramaribo, Suriname
[*]these authors contributed equally

Corresponding Author:
Raj Jadnanansing, MSc
Psychiatrisch Centrum Suriname
L. Vriesdelaan
Paramaribo
Suriname
Phone: 597 8781948
Email: raj.jadnanansing@pcs.sr

Abstract

Background: Digital health applications have been shown to be an accepted means to provide mental health information and advice in various high- and middle-income countries. Started in 2015, ehealth.sr was the first website to offer preventive information, self-tests, and unguided digital self-help for depression, anxiety symptoms, and problematic alcohol use in Suriname, an upper middle-income country in South America.

Objective: This study aimed to assess the general demographics and behavioral patterns of the visitors of ehealth.sr, as well as to evaluate different promotional channels to attract the target audience to the website.

Methods: Data collection for this study took place between August 2015 and December 2020. Conventional promotion channels such as newspaper and radio advertisements as well as social media advertisements were used to attract users to the website. The number of visits and activity on the website was registered using Google analytics and the website's internal activity log.

Results: On average, about 115 unique visitors accessed the website per month. The average number of visits to the website increased notably when social media advertisement campaigns were conducted (266 per month in 2018) compared to when traditional advertisements campaigns through papers, radio, and television were used (34 per month in 2019). Of the 1908 new visitors, 1418 (74.32%) were female. On average, visitors accessed 2 (SD 0.3) pages of the website and a session lasted 2.6 (SD 0.9) minutes. The most popular pages for intervention on the website were those for the mood or anxiety screening (731/942, 77.6%) as opposed to those for alcohol screening (211/942, 22.4%). People aged <45 years (on average, 2.2 pages per session for 3.2 minutes) made more use of the website than people aged ≥45 years (on average, 1.7 pages per session for 2 minutes).

Conclusions: Promotion via social media led to more visitors to the website than newspaper or radio advertisements. Younger age groups and females visited the website more often. The pages on preventive information and brief self-tests were visited more frequently than the self-help modules. In general, user adherence to the website in terms of the average session duration and number of viewed pages per session is low and is a key point of concern for the successful implementation of digital mental health websites.

(JMIR Form Res 2022;6(6):e33793) doi:10.2196/33793

KEYWORDS

eHealth; mental health; alcohol use disorder; depression; anxiety; Facebook; alcohol disorder; alcohol; self-help; alcoholism; Suriname

Introduction

In recent years, mental health has attracted the attention of digital technology [1] since it has the potential to transform the whole spectrum of mental health care by connecting patients, services, and health data in new ways [2]. Digital health (or eHealth as it is sometimes called) is a wide and varying concept that includes the use of communication technology for digital record keeping, web-based booking systems, web-based repeat prescriptions, and other innovative uses of technology for direct treatment [3]. The World Health Organization (WHO) defines eHealth technologies as the use of information and communication technologies for the evaluation and promotion of health [4].

The application of information and communication technologies to support national health care services is rapidly expanding and has become increasingly important globally [4]. Consequently, donor agencies and foundations invest in many digital health interventions in low- and middle-income countries (LMIC), aiming to provide a way to address major deficiencies in access to safe, effective, and affordable health services [5]. On average, nearly half the global population resides in countries where there is fewer than 1 psychiatrist per 200,000 inhabitants. This reality is reflected primarily in LMIC where, for example, up to 85% of patients with severe mental disorders do not receive treatment for their disorder [6].

The implementation of eHealth promises a number of potential benefits to the health system, including, but not limited to, increased efficiency in health care, improvement in quality, the reduction of the cost of care, and the improvement of health system governance [7]. With these aspects, the provision of health care extends beyond its conventional boundaries. In addition, the impact of eHealth is that it enables personalized mental health care throughout the health system. Furthermore, it makes mental health care available at home, work, and school and focuses on prevention, education, and self-management outside the boundaries of clinics and hospitals [4].

Due to the wide availability of information and communication technologies, LMIC have developed eHealth strategies to prevent and treat mental health disorders and substance use disorders. For instance, HDep (Help for Depression or Ayuda para depresión [ADep] in Spanish) is an open-access and free web-based, psychoeducational, and cognitive-behavioral mental health program created in Mexico with a single goal—to help recognize and prevent depression [8]. HDep has the potential to serve as a useful tool for educating people about depression in order to change negative thinking patterns and be a source of social support [8].

As a low-income country, Bangladesh received special attention in 2013 for its innovative use of eHealth, especially in a mental health setting [9]. A detailed probe of the health system concluded that Bangladesh has made enormous progress and currently has the longest life expectancy and lowest mortality rates of infants and children under 5 years in South Asia, despite spending less on health care than several neighboring countries. Much of this can be attributed to the contribution of a web-based electronic health record and mobile Health innovations [10].

The advent of eHealth could thus be a powerful strategy for the delivery of mental health services, especially in LMIC. In this paper, we focus on web-based mental health care in Suriname, an upper middle-income country. In Suriname, eHealth is made possible through the website [11] designed by Arkin Mental Health Care and the Center for Psychiatry in Suriname (PCS) and hosted by the PCS. Previous studies have found a huge gap in the treatment of mental disorders such as alcohol abuse and depression and anxiety disorders [12,13]. These studies identified a treatment gap in Suriname, a transitional country. This is the main reason why more attention has been given to a digital health intervention to address substance abuse and mental health disorders in Suriname 12,13. Currently, it is difficult for people to find specific websites on the internet and make the distinction in quality between different sites [14]. Targeted advertisement is frequently recommended for those seeking specific information and help [15].

The aim of this study was to assess the general demographics and behavioral patterns of visitors on a mental health website offering psychoeducational information, self-tests, and unguided self-help interventions developed in LMIC. In addition, we also investigated the best way we can attract the target audience to a mental health website for LMIC such as Suriname.

Methods

Setting

Suriname is a middle-income, Caribbean, and South American country with approximately 575,990 inhabitants. Web-based mental health for the Surinamese community is made possible through the eHealth.sr website [11] hosted by the PCS. This website was developed in collaboration with Arkin Mental Health Care in the context of a survey research project on the prevalence and treatment gap for substance use disorders, depression, and anxiety among the population of the Surinamese districts Nickerie and Paramaribo [12,13]. Respondents in this survey project who scored above cutoff points for alcohol abuse, depression, or anxiety were referred to the eHealth.sr website.

eHealth.sr Website

The eHealth.sr website contains modules with psychoeducational information, self-screening tests, and structured self-help for problematic alcohol use, depression, and anxiety. The website is offered in the main language of Suriname, which is Dutch. Most inhabitants of Suriname have at least a basic understanding and reading ability in Dutch.

Problematic Alcohol Use Modules

The psychoeducational information module on alcohol addresses the short- and long-term risks of drinking, including the risk of

developing alcohol use disorder and what can be done to mitigate these risks.

The self-test module is a questionnaire with feedback and includes a 7-day timeline follow-back to assess alcohol use in the past week, the Alcohol Use Disorder Identification Test, an alcohol-problems screening instrument developed by the WHO [16], and the Readiness to Change Questionnaire–Drinking [17], which is based on the Transtheoretical Model of Change [18].

After completing the questionnaires, the participants will receive automated personalized normative feedback. The self-help module contains a fully automated digital alcohol self-help program, originally developed by Jellinek [19] and not supervised by a professional. This multisession program is based on cognitive behavioral therapy and motivational interviewing (MI) techniques. The original version of this self-help intervention has shown positive results on reductions in alcohol use [20].

In this self-help intervention, various options that enable the person seeking help to determine the advantages and disadvantages of both the use as well as quitting of alcohol are offered. The participant can also set targets for themselves, visit the forum, and keep a diary. The module is an MI-based program and provides help-seeking tips and an overview of the results achieved in the self-help program. The part for problematic alcohol use has a time limit of 6 weeks, and eventually, the program fidelity and its achieved goals are displayed [11]. The module also offers the possibility for users to chat with each other and share their experiences with one another. All users are asked to create an account using their email address and a password before they can use the self-help program.

Depression and Anxiety Modules

The psychoeducational information module on depression and anxiety addresses the common symptoms of depression, stress, anxiety, and suicidal tendencies, as well as the prevention and treatment measures that can be taken to mitigate these symptoms.

The self-test module is a questionnaire with feedback and includes the Center for Epidemiologic Studies Depression (CES-D) scale [21] and the General Anxiety Disorder-7 (GAD-7) questionnaire [22]. The CES-D scale is a 20-item measure that rates the frequency of experiencing symptoms associated with depression, such as restless sleep, poor appetite, and feeling lonely, by respondents in the previous week. The GAD-7 questionnaire is a brief self-report scale that is used to identify probable cases of general anxiety disorder. It is reliable and shows good criterion, construct, factorial, and procedural validity [22]. After completing the questionnaires, the participants will receive automated personalized normative feedback.

The self-help module contains a fully automated digital alcohol self-help program based on problem-solving therapy (PST) and MI. It was inspired by the PST-MI Exercises of the Substance Use and Trauma Intervention (eSTRIVE) [23]. The findings of this study suggested that eSTRIVE appeared to be an effective brief intervention among adults presenting to emergency departments in South Africa [23].

Our self-help program offers PST-MI in 4 sessions. The participant will learn to recognize and handle problems as either unimportant or irrelevant, important but not resolvable, or important and resolvable. The goal for them is to regain control of their life, discover what is important to them, and learn how to conquer problems like nervousness, depression, or anxiety.

Procedure and Data Collection

The website [11] was launched on August 1, 2015, prior to the population survey in Nickerie (which was held from August 15-27, 2015). Interest was generated in this period during the informative sessions with the doctors, district commissioners, police officers, and civil servants that helped during the survey.

At various occasions, promotion efforts were made to attract the target audience to the website by handing out eHealth flyers at regional health care centers and through the Facebook page and newspaper advertisements. The 2015-2016 population survey in Nickerie and Paramaribo took place from July 23 to August 17, 2016, during which all respondents were informed and handed a folder about the website [11]. During these months, extra nationwide promotion took place via radio interviews and newspaper advertisements. In August 2018, we created a Facebook page and used Facebook advertisements to attract the target audience to the website. Throughout the whole year in 2019, we advertised the existence of the website once a week in daily newspapers as well as on several occasions on television and radio programs. Accordingly, we used the data obtained in 2016 and 2019 to assess the effects of traditional media and the data obtained around August 2018 to assess the effect of promotion through social media.

Page views and visitors' behavior were recorded anonymously using Google Analytics. The responses of the participants in the self-test were also recorded anonymously to give them automated normative feedback. "Users" refers to the number of unique visitors to the site, whereas "sessions" refers to the total number of visits to the site, which includes both new and repeat visits [24]. Data in the self-help program were not anonymous, and as no informed consent was sought from the self-help participants, we cannot report the data collected in the self-help programs. To present data on the use patterns of the website modules, we used descriptive statistics.

Ethical Considerations

As no data were prospectively collected solely for the purpose of this study and no participants were subjected to research procedures, this study was deemed exempt from medical ethics approval. The website is a result of a larger study that was approved by the Ministry of Health of Suriname (CMWO201504). Due to the anonymity of the persons entering their data on the site, consent was not required. In addition, consent was not obtained to guarantee the anonymity of the participants, thus they were prevented from entering personal data on any occasion.

Results

The average number of monthly visits to the eHealth website from 2015 to 2020 were 202 per month in 2015, 94 per month in 2016, 39 per month in 2017, 266 per month in 2018, 34 per

month in 2019, and 59 per month in 2020. On average, about 115 unique visitors accessed the website per month. The value for 2015 represents only 5 months because the program started in August. We saw a relatively high monthly number of visits in 2015 and 2016 and then again in 2018, with a small peak in 2020. The high number of visits of 2015 and 2016 were probably due to the introduction of the website to individuals during the survey. In 2018, there was an active advertisement through the Facebook page of the institute. A systematic review recommended this type of advertising as being extremely successful [25]. The number of visits to the website increased notably when social media advertisement campaigns was conducted (with a peak of 2809 in August 2018) compared to when traditional advertisements campaigns through papers, radio, and television were used (34 per month in 2019). In 2020, COVID-19 limited in-person visits to the clinics because all care was downgraded to emergency cases only.

Table 1 shows the gender distribution of the website visits. More female than male visitors made use of the website—1418 of the 1908 (74%) new visitors were female. Table 2 represents the age distribution of the sessions. We observed that the number of sessions increased with age.

Figure 1 shows the fraction of new sessions compared to the total sessions distributed by age, whereas Figure 2 shows the bounce rate (the percentage of sessions with only 1 page visited). On average, visitors accessed 2 (SD 0.3) pages of the website and a session lasted 2.6 (SD 0.9) minutes. The most popular pages on intervention modules on the website were those for the mood or anxiety screening (731/942, 77.6%) as opposed to those for alcohol screening (211/942, 22.4%). People aged <45 years (on average, 2.2 pages per session for 3.2 minutes) made more use of the website than people aged ≥45 years (on average, 1.7 pages per session for 2 minutes).

Table 3 shows the average time spent on the site. There were 2 peaks: the initial 10 seconds spent at the pages and sessions taking longer than 3 minutes. Thus, the majority of the participants decided within 10 seconds to quit or to continue for longer than 3 minutes.

Table 4 gives an overview of the numbers of participants starting the screening program for alcohol and mood or anxiety disorders and finishing the last page of the module. The number of participants finishing the modules for mood or anxiety disorders was more than those for possible problems with alcohol (odds ratio 2.137, 95% CI 1.583-2.888).

Table 5 presents the number of participants who were invited to enter the intervention for self-help for problem drinkers and mood or anxiety symptoms (eSTRIVE). The follow-up invitation was given 1 month after the initial invitation, while the participants were reminded after 1 week to finish the module. The number of participants of the mood or anxiety symptoms module (n=165) was notably higher than the number of participants of the problem drinking module (n=20). Overall, compared to the number of website visitors, the uptake of the self-help modules was low.

Table 1. Gender distribution of website visits.

Gender	New user (n=1908), n (%)	Session (n=2605), n (%)
Female	1418 (74.32)	1951 (74.89)
Male	490 (25.68)	654 (25.11)

Table 2. Session characteristics by age group distribution.

Age group (years)	Number of sessions (N=2528), n (%)	Pages viewed per session, mean (SD)	Average session length (s), mean (SD)
18-24	291 (11.51)	2.31 (0.4)	237.24 (92)
25-34	443 (17.52)	2.18 (0.3)	192.53 (75)
35-44	337 (13.33)	2.09 (0.3)	140.55 (55)
45-54	374 (14.79)	1.8 (0.3)	125.03 (49)
55-64	535 (21.16)	1.89 (0.3)	138.59 (54)
≥65	548 (21.68)	1.55 (0.2)	93.19 (36)

Figure 1. Rate of new sessions per age group.

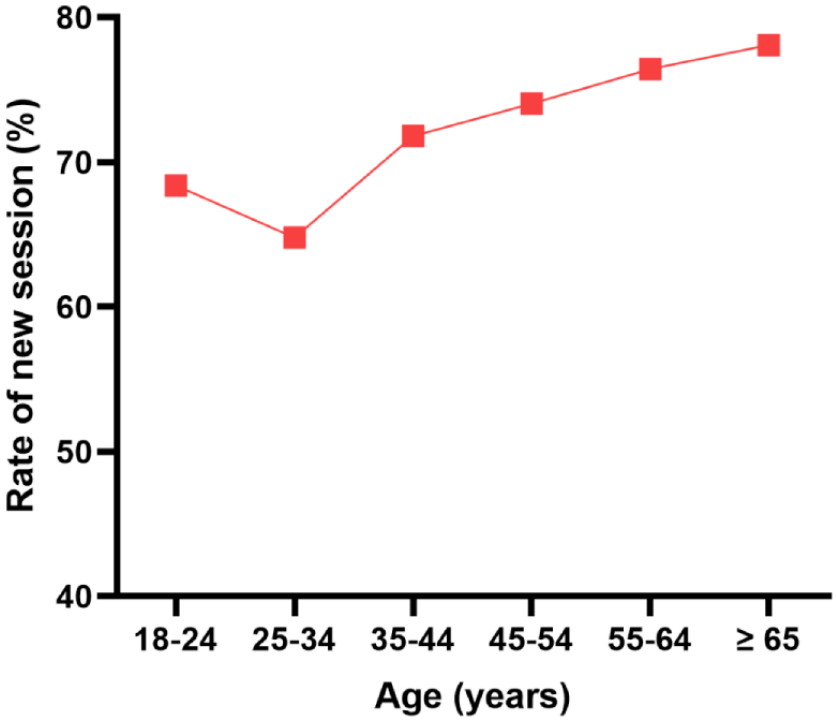


Figure 2. Bounce rate per age group.

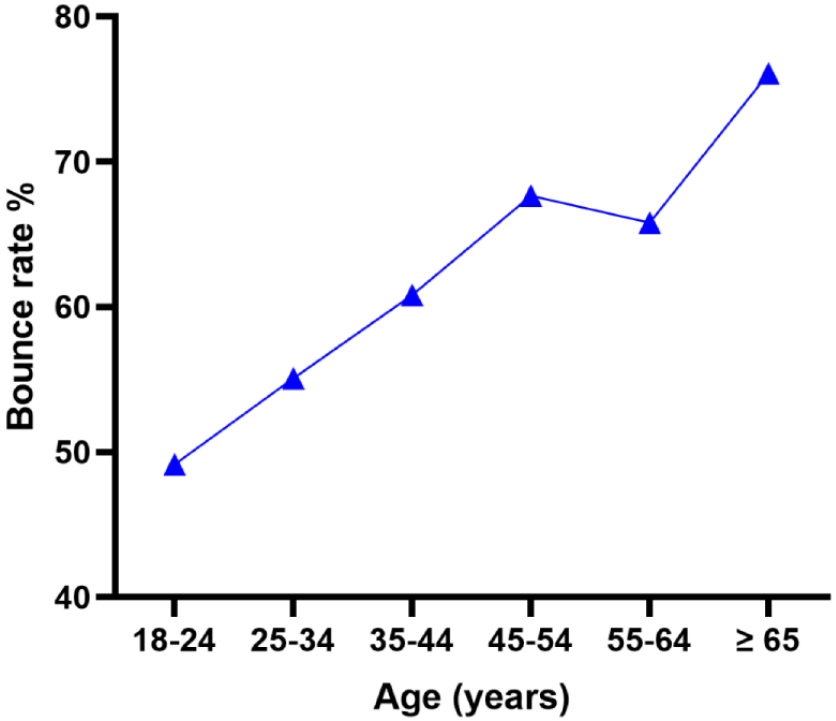


Table 3. The frequency of average time in seconds spent on the site.

Length of session (seconds)	Number of sessions (n=8973), n (%)	Number of pages displayed (n=14,914), n (%)
0-10	6888 (76.76)	7470 (50.09)
11-30	385 (4.29)	949 (6.36)
31-60	272 (3.03)	686 (4.6)
61-180	311 (3.47)	1017 (6.82)
181-600	642 (7.15)	2337 (15.67)
601-1800	421 (4.69)	1998 (13.4)
>1800	54 (0.6)	457 (3.06)

Table 4. The number of participants starting and finishing the screening program for alcohol and mood or anxiety disorders.

Screening	Participants entering the site (n=942), n (%)	Participants finishing the last page (n=521), n (%)
Alcohol	211 (22.3)	62 (11.9)
Mood or anxiety	731 (77.6)	459 (88.1)

Table 5. The number of participants invited to enter the intervention for self-help with suspected alcoholism and mood or anxiety disorders (eSTRIVE[a]).

Module	Invited	Reminded	Completed
Alcohol self-help			
Baseline (n=20), n (%)	20 (100)	18 (90)	4 (20)
Follow-up (n=20), n (%)	20 (100)	20 (100)	1 (5)
Mood or anxiety disorders (eSTRIVE)			
Baseline (n=165), n (%)	165 (100)	154 (93.3)	19 (11.5)
Follow-up (n=159), n (%)	159 (100)	157 (98.7)	7 (4.4)

[a]eSTRIVE: Exercises of the Substance Use and Trauma Intervention.

Discussion

Principal Findings

This study aimed to describe the general demographics and behavioral patterns of the visitors of an eHealth website provided by the PCS and the impact of the different forms of promotion on this site. We found that younger age groups were more likely to use the site for diagnostic and intervention purposes. There were more female than male visitors to the website. Furthermore, we observed that the eHealth module parts that were less intensive got more visitors than those that were more intensive since the launch in 2015. The number of visitors of the eSTRIVE module for depression were also higher than the module for problematic alcohol use. Another important finding is that whenever there was an active or intensive campaign in person or through social media, the number of visitors increased. Regular advertisement in newspapers and other nondigital media did not result in the same number of visits. Additionally, there was a small peak in 2020, which is probably ascribable to the measures taken with COVID-19, where clinics were closed and only used for emergency cases.

Comparison With Prior Work

The decline of internet use for eHealth with age that we observed is very consistent with findings in high-income countries. A population-based survey from 2005-2007 in several European countries showed that the use of internet modules for general health was highest among people aged 18-25 years (70% to 80%) and declined progressively to under 20% among people aged >65 years [26]. Another survey in 2012 in the United States showed that the odds ratio for someone in the age group of 18-34 years who looked for health information on the internet to self-help was 3.51 (95% CI 1.66-7.44) compared to someone from the age group of >65 years [26]. In contrast to our findings, none of the abovementioned studies found a difference in gender use. However, a more recent Spanish study found that women used the modules more than men [27]. In accordance with our findings, the Spanish study also found that the modules for mood disorders were highest in use among the other modules [27].

Comparable studies in LMIC are not available. To our knowledge, this study is the first one analyzing these data in LMIC. The study shows that with proper advertisement it is possible to apply eHealth modules for mental disorders in LMIC. This is favorable since recent surveys have shown a significantly high treatment gap for alcohol abuse disorders [12] and depression and anxiety [13]. The program fully adheres to the WHO Mental Health Gap Action Program. First, it stimulates the potential patients to look for help. Second, it brings the patients closer to the availability of mental health professionals, and in this way also calls for the commitment of these professionals as stated in the goals of the WHO program.

Finally, we found that personal promotion and advertisements through social media resulted in the largest increase of visitors to the website. Traditional promotion through advertisements in the local newspapers and conventional media like radio and television did not contribute to a higher number of visits to the site. In general, peers and facilitators may increase the success rate of remote health programs [28], and this can be applied to promotion through social media [25].

Limitations

A limitation of the program is that the promotion was initially in only 2 regions of the country, which could bias the number of visits to the website. However, more than 75% of the population resides in the capital, which minimizes the amount of expected bias. Another limitation is that the module is offered only in Dutch, thereby excluding those who do not use the language. Furthermore, the requirement of decent internet to complete the survey and then enter a possible treatment module limits the availability for the entire population. Due to the anonymous nature of the website, we were unable to assess the real-world impact of these limitations.

Conclusion

In summary, we found that younger age groups and females visited the website more often. Furthermore, there was a success rate of at least 25% for both the alcohol abuse and depression or anxiety modules. Finally, we found that advertising through Facebook had the best effect on the use of these modules. We recommend that the use of an eHealth mental health intervention module is feasible to bridge part of the treatment gap in LMIC such as Suriname, if it is promoted adequately. Furthermore, to increase the use of the facility, it would be necessary to promote this kind of intervention among the older adults and men. We also recommend that the best way for promotion is through the personal involvement of doctors, nurses, and students. Advertisement through social media seems to work the best in these settings. Since the use of smartphones nowadays is more acceptable and widespread than desktop and laptop devices, it is recommended that future research develop a user-friendly app for smartphones.

Conflicts of Interest

None declared.

References

1. Aguilera A. Digital technology and mental health interventions: opportunities and challenges. Arbor 2015 May 12;191(771):a210. [doi: 10.3989/arbor.2015.771n1012]
2. Hollis C, Morriss R, Martin J, Amani S, Cotton R, Denis M, et al. Technological innovations in mental healthcare: harnessing the digital revolution. Br J Psychiatry 2015 Apr;206(4):263-265. [doi: 10.1192/bjp.bp.113.142612] [Medline: 25833865]
3. Oh H, Rizo C, Enkin M, Jadad A. What is eHealth?: a systematic review of published definitions. World Hosp Health Serv 2005;41(1):32-40. [Medline: 15881824]
4. National eHealth strategy toolkit. World Health Organization. 2012. URL: https://apps.who.int/iris/bitstream/handle/10665/75211/9789241548465_eng.pdf [accessed 2022-05-27]
5. Adebayo AK, Bankole FA, Adebayo JA. The impact and challenges of information and communication technology of health care delivery in Nigeria. Madonna Univ Niger Fac Law Law J 2021;6(2):34-45 [FREE Full text]
6. Dixon A, McDaid D, Knapp M, Curran C. Financing mental health services in low- and middle-income countries. Health Policy Plan 2006 May;21(3):171-182. [doi: 10.1093/heapol/czl004] [Medline: 16533860]
7. Eysenbach G. What is e-health? J Med Internet Res 2001;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
8. Lara MA, Tiburcio M, Aguilar Abrego A, Sánchez-Solís A. A four-year experience with a web-based self-help intervention for depressive symptoms in Mexico. Rev Panam Salud Publica 2014;35(5-6):399-406. [Medline: 25211568]
9. Balabanova D, Mills A, Conteh L, Akkazieva B, Banteyerga H, Dash U, et al. Good Health at Low Cost 25 years on: lessons for the future of health systems strengthening. Lancet 2013 Jun 15;381(9883):2118-2133. [doi: 10.1016/S0140-6736(12)62000-5] [Medline: 23574803]
10. Adams AM, Ahmed T, El Arifeen S, Evans TG, Huda T, Reichenbach L, Lancet Bangladesh Team. Innovation for universal health coverage in Bangladesh: a call to action. Lancet 2013 Dec 21;382(9910):2104-2111. [doi: 10.1016/S0140-6736(13)62150-9] [Medline: 24268605]
11. Mogelijk gemaakt door Psychiatrisch Centrum Suriname en Arkin. Ehealth.sr. URL: https://data.ehealth.sr/Home/ [accessed 2021-07-26]
12. Jadnanansing R, Blankers M, Dwarkasing R, Etwaroo K, Lumsden V, Dekker J, et al. Prevalence of substance use disorders in an urban and a rural area in Suriname. Trop Med Health 2021 Feb 02;49(1):12 [FREE Full text] [doi: 10.1186/s41182-021-00301-7] [Medline: 33526098]
13. Jadnanansing R, de Beurs E, Etwaroo K, Blankers M, Dwarkasing R, Peen J, et al. A survey of depression and anxiety disorders in urban and rural Suriname. BMC Public Health 2022 Jan 08;22(1):51 [FREE Full text] [doi: 10.1186/s12889-021-12454-5] [Medline: 34998381]

14. Levy S, Gvili Y. How credible is e-word of mouth across digital-marketing channels?: the roles of social capital, information richness, and interactivity. J Advert Res 2015 Mar 01;55(1):95-109. [doi: 10.2501/jar-55-1-095-109]
15. Singh K, Drouin K, Newmark LP, Lee J, Faxvaag A, Rozenblum R, et al. Many mobile health apps target high-need, high-cost populations, but gaps remain. Health Aff (Millwood) 2016 Dec 01;35(12):2310-2318. [doi: 10.1377/hlthaff.2016.0578] [Medline: 27920321]
16. Johnson JA, Lee A, Vinson D, Seale JP. Use of AUDIT-based measures to identify unhealthy alcohol use and alcohol dependence in primary care: a validation study. Alcohol Clin Exp Res 2013 Jan;37 Suppl 1:E253-E259. [doi: 10.1111/j.1530-0277.2012.01898.x] [Medline: 22834916]
17. Heather N, Rollnick S, Bell A. Predictive validity of the Readiness to Change Questionnaire. Addiction 1993 Dec;88(12):1667-1677. [doi: 10.1111/j.1360-0443.1993.tb02042.x] [Medline: 8130706]
18. DiClemente CC, Prochaska JO. Toward a comprehensive, transtheoretical model of change: stages of change and addictive behaviors. In: Miller WR, Heather N, editors. Treating Addictive Behaviors. New York, NY: Plenum Press; 1998:3-24.
19. Jellinek: de expert op het gebied van verslavingszorg. Jellinek. URL: https://www.jellinek.nl/ [accessed 2021-09-03]
20. Blankers M, Koeter MWJ, Schippers GM. Internet therapy versus internet self-help versus no treatment for problematic alcohol use: a randomized controlled trial. J Consult Clin Psychol 2011 Jun;79(3):330-341. [doi: 10.1037/a0023498] [Medline: 21534652]
21. Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. App Psychol Meas 1977 Jun 01;1(3):385-401. [doi: 10.1177/014662167700100306]
22. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
23. Sorsdahl K, Stein DJ, Corrigall J, Cuijpers P, Smits N, Naledi T, et al. The efficacy of a blended motivational interviewing and problem solving therapy intervention to reduce substance use among patients presenting for emergency services in South Africa: a randomized controlled trial. Subst Abuse Treat Prev Policy 2015 Nov 14;10:46 [FREE Full text] [doi: 10.1186/s13011-015-0042-1] [Medline: 26576946]
24. Hanlon C. Users vs. sessions: what's the difference? Whole Whale. URL: https://www.wholewhale.com/tips/users-vs-sessions-whats-difference/ [accessed 2021-09-05]
25. Whitaker C, Stevelink S, Fear N. The use of Facebook in recruiting participants for health research purposes: a systematic review. J Med Internet Res 2017 Aug 28;19(8):e290 [FREE Full text] [doi: 10.2196/jmir.7071] [Medline: 28851679]
26. Kummervold PE, Chronaki CE, Lausen B, Prokosch HU, Rasmussen J, Santana S, et al. eHealth trends in Europe 2005-2007: a population-based survey. J Med Internet Res 2008 Nov 17;10(4):e42 [FREE Full text] [doi: 10.2196/jmir.1023] [Medline: 19017584]
27. Lopez-Castroman J, Abad-Tortosa D, Cobo Aguilera A, Courtet P, Barrigón ML, Artés A, et al. Psychiatric profiles of eHealth users evaluated using data mining techniques: cohort study. JMIR Ment Health 2021 Jan 20;8(1):e17116 [FREE Full text] [doi: 10.2196/17116] [Medline: 33470943]
28. Kim H, Ray CD, Veluscek AM. Complementary support from facilitators and peers for promoting mHealth engagement and weight loss. J Health Commun 2017 Nov 10;22(11):905-912. [doi: 10.1080/10810730.2017.1373876] [Medline: 29125393]

Abbreviations

CES-D: Center for Epidemiologic Studies Depression
eSTRIVE: Exercises of the Substance Use and Trauma Intervention
GAD-7: General Anxiety Disorder-7
LMIC: low- and middle-income countries
MI: motivational interviewing
PCS: Center for Psychiatry in Suriname
PST: problem-solving therapy
WHO: World Health Organization

Edited by A Mavragani; submitted 23.09.21; peer-reviewed by A Garcia, R De Boni; comments to author 21.01.22; revised version received 08.04.22; accepted 22.05.22; published 09.06.22.

Please cite as:
Jadnanansing R, Dekker J, Etwaroo K, Dwarkasing R, Lumsden V, Bipat R, Blankers M
General Demographics and Behavioral Patterns of Visitors Using a Self-help Website for Identification of and Intervention in Alcoholism and Common Mental Disorders in Suriname: Descriptive Study
JMIR Form Res 2022;6(6):e33793
URL: https://formative.jmir.org/2022/6/e33793
doi:10.2196/33793
PMID:35679108

Original Paper

Assessing the Usability of a Novel Wearable Remote Patient Monitoring Device for the Early Detection of In-Hospital Patient Deterioration: Observational Study

Edward Itelman[1], MD; Gadi Shlomai[1], MD; Avshalom Leibowitz[1], MD; Shiri Weinstein[1], MD; Maya Yakir[1], MD; Idan Tamir[1], MD; Michal Sagiv[1], MD; Aia Muhsen[1], MD; Maxim Perelman[1], MD; Daniella Kant[1], MD; Eyal Zilber[1], MD; Gad Segal[1], Prof Dr

Chaim Sheba Medical Center, Sheba Beyond, Virtual Hospital, Ramat Gan, Israel

Corresponding Author:
Gad Segal, Prof Dr
Chaim Sheba Medical Center
Sheba Beyond
Virtual Hospital
Sheba Road 2
Ramat Gan, 555710
Israel
Phone: 972 526669580
Email: gad.segal@sheba.health.gov.il

Abstract

Background: Patients admitted to general wards are inherently at risk of deterioration. Thus, tools that can provide early detection of deterioration may be lifesaving. Frequent remote patient monitoring (RPM) has the potential to allow such early detection, leading to a timely intervention by health care providers.

Objective: This study aimed to assess the potential of a novel wearable RPM device to provide timely alerts in patients at high risk for deterioration.

Methods: This prospective observational study was conducted in two general wards of a large tertiary medical center. Patients determined to be at high risk to deteriorate upon admission and assigned to a telemetry bed were included. On top of the standard monitoring equipment, a wearable monitor was attached to each patient, and monitoring was conducted in parallel. The data gathered by the wearable monitors were analyzed retrospectively, with the medical staff being blinded to them in real time. Several early warning scores of the risk for deterioration were used, all calculated from frequent data collected by the wearable RPM device: these included (1) the National Early Warning Score (NEWS), (2) Airway, Breathing, Circulation, Neurology, and Other (ABCNO) score, and (3) deterioration criteria defined by the clinical team as a “wish list” score. In all three systems, the risk scores were calculated every 5 minutes using the data frequently collected by the wearable RPM device. Data generated by the early warning scores were compared with those obtained from the clinical records of actual deterioration among these patients.

Results: In total, 410 patients were recruited and 217 were included in the final analysis. The median age was 71 (IQR 62-78) years and 130 (59.9%) of them were male. Actual clinical deterioration occurred in 24 patients. The NEWS indicated high alert in 16 of these 24 (67%) patients, preceding actual clinical deterioration by 29 hours on average. The ABCNO score indicated high alert in 18 (75%) of these patients, preceding actual clinical deterioration by 38 hours on average. Early warning based on wish list scoring criteria was observed for all 24 patients 40 hours on average before clinical deterioration was detected by the medical staff. Importantly, early warning based on the wish list scoring criteria was also observed among all other patients who did not deteriorate.

Conclusions: Frequent remote patient monitoring has the potential for early detection of a high risk to deteriorate among hospitalized patients, using both grouped signal-based scores and algorithm-based prediction. In this study, we show the ability to formulate scores for early warning by using RPM. Nevertheless, early warning scores compiled on the basis of these data failed to deliver reasonable specificity. Further efforts should be directed at improving the specificity and sensitivity of such tools.

Trial Registration: ClinicalTrials.gov NCT04220359; https://clinicaltrials.gov/ct2/show/NCT04220359

(JMIR Form Res 2022;6(6):e36066) doi:10.2196/36066

KEYWORDS

remote patient monitoring; noninvasive monitoring; general ward; early warning score system; patient deterioration; clinical prediction; wearable devices; uHealth

Introduction

Validated tools for the early identification of high risk for clinical deterioration in hospitalized patients, or early warning score (EWS) systems, would be of high medical value in clinical practice. A large meta-analysis of previous studies tried to evaluate the effectiveness of rapid response teams for the reduction of in-hospital death in such circumstances [1]. The analysis failed to reach firm conclusions owing to the low quality of design and subsequent outcomes of such studies. One potential cause could be the fact that in different clinical scenarios (eg, general wards vs surgical departments) the clinical circumstances, the classifications used to define the clinical deterioration, and the competencies of the clinical staff are heterogenous [2,3]. Therefore, ideally, early detection technologies and applied prediction algorithms should be tailored for specific patient populations and clinical scenarios.

A common method, used worldwide in general-internal medicine departments, for the early identification of deterioration is placing the patient in a telemetry bed. A retrospective analysis of the effectiveness and potential abuse of this method found that when analyzed retrospectively, only one-quarter of telemetry days during hospitalization were deemed appropriate [4]. Moreover, they described that eliminating unnecessary telemetry days would result in significant cost saving. Interestingly, they did not find any cases of deterioration among patients who were not connected to telemetry devices. This shows that the medical staff was highly professional yet too sensitive, having admitted patients to the telemetry bed frequently.

Possible ways to generate an early risk identification flag is to rely on automatically grouped physiological signals incorporated into different scoring systems, or using artificial intelligence (AI)-based computerized algorithms, rather than counting on follow-up observations by professional staff members. For example, a multicenter retrospective analysis of electronic health records' data from all patients admitted to 5 US hospitals during the years 2008-2013 showed that prediction of the composite outcome of in-hospital cardiac arrest, the need for intensive care unit (ICU) transfer, and death within 24 hours of observation were higher when conducted using a computerized score [5]. This was also found in the setting of the high-acuity area of an emergency department [6]. Nevertheless, training and competency of professional staff members are key components in every program intended to assimilate computerized predictive tools in hospital departments [7].

Unlike the standard spot-check vital sign measurements that are conducted over a short period and could miss changes in parameters, frequent and automated vital sign collection for longer periods using remote patient monitoring (RPM) platforms with data transmission into algorithm-based computerized systems will potentially be better equipped to detect early changes and alerts of various risks [8-11]. It is also accepted that such systems would be of benefit if they are simple and easy to use, frequently measure multiple vital signs, incorporated into the workflow of the health care providers, improve patient outcomes, and would be of help to the medical staff in addition to other measures to improve patient surveillance [11-14]. We expect monitoring systems and early warning scores to be sensitive and specific. The currently used EWS systems, which are collected infrequently by the medical teams, are known to have a relatively high sensitivity and low specificity [15].

This study aimed to assess whether frequent RPM has the potential for early detection of the risk to deteriorate, using grouped signal-based scores, compared to clinical detection of deterioration by the medical staff.

Methods

Study Design and Overview

This prospective observational clinical study with retrospective analysis of the data was conducted in 2 general wards of a large tertiary medical center. Patients determined to be at high risk to deteriorate upon admission and assigned to a telemetry bed were included, after signing an informed consent form. On top of the standard telemetry overhead monitoring devices used in the general ward (Mindray; e PM 10M), a wireless, wearable monitor was attached to the chest of each patient, and monitoring was conducted in parallel. Multimedia Appendix 1 shows a CONSORT (Consolidated Standards of Reporting Trials) flowchart of the study.

Medical treatment was provided on the basis of standard monitoring system only, as the data gathered by the wearable monitor were analyzed retrospectively, with the medical staff being blinded to it in real time. The physiological data from the wearable monitors were collected automatically every 5 minutes during the first 72 hours from admission, with no personally identifiable information besides serial numbers of the devices. Inclusion criteria were adults (aged >18 years) transferred from the emergency department and admitted to the general wards, who were determined to be at an increased risk for clinical and physiological deterioration during the first 72 hours from admission by the attending physicians (eg, patients who were hemodynamically or respiratory unstable in the emergency department, patients suspected of arrhythmia or acute coronary syndrome, and those suspected with infection and signs of sepsis). Exclusion criteria were lack of informed consent, physicians' assessment that patients will not stay in the general ward for the entirety of the first 72 hours, and technical inability to attach the chest monitor to the patients. Furthermore, patients already defined as necessitating lifesaving procedures were not included in the study.

Study Setting

Designated communication routers were deployed and installed in the departments to ensure continuous monitoring, data transmission, and automatic data collection of all measurements. Data were transferred through Bluetooth from the devices to the routers and through Wi-Fi from the routers to a data cloud. Data gathered during the first 72 hours post admission, including physiological parameters measured by the wearable monitors and clinical data and vital sign data collected using the standard-of-care devices and documented in the electronic medical record system, were retrospectively analyzed at the end of the collection phase.

Early Warning of the Risk for Deterioration

Several EWSs of the risk for deterioration were used. These included (1) the National Early Warning Score (NEWS, described in Multimedia Appendix 2) [16-19]; (2) Airway, Breathing, Circulation, Neurology, and Other (ABCNO) score (described in Multimedia Appendix 3) [20]; and (3) deterioration criteria defined by the clinical team as what they expect to have from a device providing continuous monitoring (a "wish list" score, described in Multimedia Appendix 4). In all three scoring systems, the risk scores were calculated every 5 minutes using the frequent data collected by the wearable devices. Data generated by these early warning scores were compared with those obtained from the clinical records of actual deterioration among these patients. Actual clinical deterioration of patients was defined by the medical staff as (1) needing cardiopulmonary resuscitation, (2) needing to be transferred to the ICU, (3) dead or dying, or (4) deteriorating as defined by the ABCNO criteria, relying on measurements from currently used devices in the wards and without using data derived from the wearable monitors.

The Wearable Monitoring Platform

Frequent monitoring was achieved using a wireless, noninvasive, wearable reflective photoplethysmography-based sensor (BB-613WP, Biobeat Technologies Ltd). The data were automatically transmitted immediately upon capture to a cloud-based web platform repository. Patients' values recorded every 5 minutes included 13 physiological parameters, including heart rate, blood oxygen saturation, respiratory rate, cuffless blood pressure, stroke volume, cardiac output, cardiac index, systemic vascular resistance, heart rate variability, pulse pressure, mean arterial pressure, temperature, and single-channel electrocardiograms [21-25].

Statistical and Data Analysis

We compared various EWS systems using the data collected via the continuous wearable monitoring system, with the exact time as recorded in the electronic medical record, where patients deteriorated, as detected by the medical teams.

Baseline physical parameters were calculated by averaging the first 12 measurements. Continuous data are expressed as mean (SD) values if normally distributed or median (IQR) values if skewed. Categorical variables are presented as frequency (%) values. Between-group comparisons of numerical values were carried out using an independent samples *t* test, followed by the Levene test for equality of variances. Chi-square and the Fisher exact test were used for between-group comparisons of categorial parameters. Early warning based on NEWS was defined as the initial time point in which the score was above 5. Early warning based on the ABCNO score and the local medical staff deterioration ("wish list") criteria was defined as being detected when two consecutive measurements were above or below the defined thresholds (see Multimedia Appendices 3 and 4). Patients were included in the final analysis if more than 200 sessions of measurements (each session includes 13 physiological parameters and an EWS score) were carried out per patient, which was considered the minimal volume of data to analyze deterioration during the monitoring period. We did not treat any missing data; once patients had more than 200 measurement sessions, they were considered eligible for inclusion and further analysis.

The actual clinical deterioration events detected by the medical teams on site were collected, and the sensitivity and specificity of early warning by the wearable monitoring platform based on the 3 approaches—NEWS, ABCNO score, and "wish list" criteria of changes in parameters—were assessed post hoc, relying on the combination of the documented events in the electronic medical records of the patients and the physiological data collected by the wearable monitoring system. The investigators who made these post hoc assessments were blinded to the clinical outcomes of participating patients. Another element assessed was the warning time defined as the difference (in hours) from the early detection by any of these approaches to the actual clinical detection of deterioration as documented by the medical staff. All descriptive statistical analyses were performed using SPSS (version 25; IBM Corp).

Within the context of this study, readings regarded as "spurious readings" (basic definitions of either bad signals or signals defined as out of the sensor's measurement range) were automatically removed by the monitoring platform's algorithm and not included in the analysis. Thus, all collected measurements were regarded as valid. The next step was to aggregate the 15-minute data (using all data points) into hourly measurement aggregates using Python's data analysis library [26] and to match the data with the clinical data for each subject, considering also their demographic characteristics and their medical history.

Ethics Approval

This study was approved by the institutional review board of the Sheba Medical Center, Israel (MOH_2020-07-12_009133).

Results

A total of 410 patients, fulfilling the preliminary definition by the attending physicians as being at a high risk to deteriorate during the first 72 hours after admission, were initially recruited. The median patient age was 71 (IQR 62-78) years. Of the recruited 410 patients, 217 had undergone more than 200 measurement sessions using the wearable monitors during their hospitalization (average monitoring time 48 hours, range 25-131 hours) and thus were included in the final analysis. In total, 13 parameters were collected within each measurement session, resulting in approximately 3700 measurements per day per for

each of the 217 patients included in the final analysis. When considering at least 2 days of the monitoring period, the number of data points collected during the study crossed 2,000,000 altogether. Of the 217 patients, 130 (59.9%) were male. Demographic details of the participants (upon admission) are provided in Table 1.

Baseline measurements were not significantly different between patients who deteriorated and those who did not. Actual clinical deterioration was detected by the medical staff in 24 of the 217 (11.1%) patients (Table 2).

When analyzing the frequent data collected by the wearable monitors, the NEWS method provided an early warning in 16 of the 24 (67%) patients who deteriorated at 29 hours on average before actual deterioration was detected by the medical staff (an example from one patient is shown in Figure 1). The ABCNO criteria were met in 18 of the 24 (75%) patients at 38 hours on average before actual deterioration was detected by the medical staff. Early warning based on the "wish list" criteria was detected in all 24 patients at 40 hours on average before it was detected by the medical staff.

In total, 193 patients did not experience clinical deterioration during the index hospitalization. However, NEWS provided early warning alerts in 150 of the 193 (77.7%) patients, ABCNO criteria were met in 162 of the 193 (83.9%) patients, and when following the "wish list" criteria, all 193 patients who did not deteriorate had early warning alerts.

When measuring the sensitivity and specificity of the methods applied, NEWS revealed a sensitivity of 67% and specificity of 22%; ABCNO score, 75% and 16%; and the "wish list" criteria, 100% and 0%, respectively (Table 3).

Table 1. Demographic data of study participants with >200 measurements upon admission (N=217).

Characteristics	No deterioration (n=193)	Deteriorated (n=24)	*P* value
Age (years), mean (SD)	70.3 (15.2)	71.8 (13.1)	.65
Sex (male/female), n/n	115/79	16/8	.52
BMI (kg/m^2), mean (SD)	27.1 (5.4)	24.8 (6.2)	.05
Ethnicity, n			.82
Ashkenazy	83	11	
Sephardi	104	13	
Arabic	6	0	
Other	1	0	
Blood oxygen saturation (%), mean (SD)	92.9 (10.9)	95.5 (2.2)	.30
Respiratory rate (breaths/min), mean (SD)	17.7 (3.5)	17.2 (2.2)	.50
Temperature (°C), mean (SD)	37.3 (0.6)	37.2 (0.6)	.31
Heart rate (beats/min), mean (SD)	80.9 (17.7)	79.9 (17.4)	.80
Systolic blood pressure (mm Hg), mean (SD)	129.3 (24.4)	131.0 (25.1)	.75
Diastolic blood pressure (mm Hg), mean (SD)	68.7 (14.0)	72.2 (14.3)	.24
Pulse pressure (mm Hg), mean (SD)	60.6 (20.9)	58.8 (24.3)	.69
Mean arterial pressure (mm Hg), mean (SD)	88.9 (15.2)	91.8 (14.7)	.37
Stroke volume (mL), mean (SD)	72.0 (13.6)	74.7 (17.0)	.37
Cardiac output (L/min), mean (SD)	5.7 (0.9)	5.8 (0.9)	.39
Cardiac index (L/min/m^2), mean (SD)	3.1 (0.6)	3.1 (0.9)	.49
Systemic vascular resistance (dynes•s/cm^5), mean (SD)	1283.0 (256.1)	1305.3 (319.9)	.70
Background diagnosis, n (%)			
Ischemic heart disease	61 (31.6)	11 (46)	.17
Hypertension	119 (73)	17 (71)	.50
Congestive heart failure	21 (10.9)	4 (17)	.49
Diabetes mellitus	77 (39.9)	9 (38)	>.99
Obesity	21 (10.9)	0 (0)	.14
Valve disease	15 (7.8)	2 (8)	>.99
Chronic obstructive pulmonary disease	38 (19.7)	2 (8)	.26
Asthma	12 (6.2)	2 (8)	.66
Cerebrovascular accident	26 (13.5)	3 (13)	>.99
Chronic kidney disease	48 (24.9)	6 (25)	>.99
Epilepsy	2 (1.0)	0 (0)	>.99
Previous surgery	107 (55.4)	12 (50)	.67
Arrhythmia	60 (31.1)	8 (33)	.82
Anemia	27 (14.0)	3 (13)	>.99
Active malignancy	40 (20.7)	6 (25)	.60
Past malignancy	11 (5.7)	3 (13)	.19
Thyroid	27 (14.0)	4 (17)	.76
Pacemaker	13 (6.7)	1 (4)	>.99
Depression	8 (4.1)	0 (0)	.60
Bronchiectasis or cystic fibrosis	2 (1.0)	2 (8)	>.99

Characteristics	No deterioration (n=193)	Deteriorated (n=24)	*P* value
COVID-19	2 (1.0)	1 (4)	.30
Length of stay (days), mean (SD)	2.2 (2.3)	2.4 (0.8)	.61

Table 2. Comparison of different tools for early detection of patient deterioration (relating to first-time alerts only).

Tools	Patients in whom an early risk alert was generated by the scoring system, n (%)		Time of detection prior to actual clinical deterioration (hours), n
	No deterioration (n=193)	Deteriorated (n=24)	
National Early Warning Score	150 (77.7)	16 (67)	29
Airway, Breathing, Circulation, Neurology, and Other score[a]	162 (83.9)	18 (75)	38
The clinical definition of deterioration by local medical staff (“wish list”)	193 (100)	24 (100)	40

[a]A locally implemented version of the Airway, Breathing, Circulation, Disability, Exposure criteria for identification of patients' deterioration, as described in Multimedia Appendix 3.

Figure 1. Trends of continuous data gathered by the monitoring platform. Sample of the monitoring data from a single patient, showing systolic blood pressure (mm Hg), heart rate (beats/min), respiratory rate (breaths/min), blood oxygen saturation (%), and markings of warnings and prediction. The black line indicates the time of actual clinical detection of deterioration by the medical staff. Red lines indicate times at which high-risk warnings were provided by the platform using the National Early Warning Score.

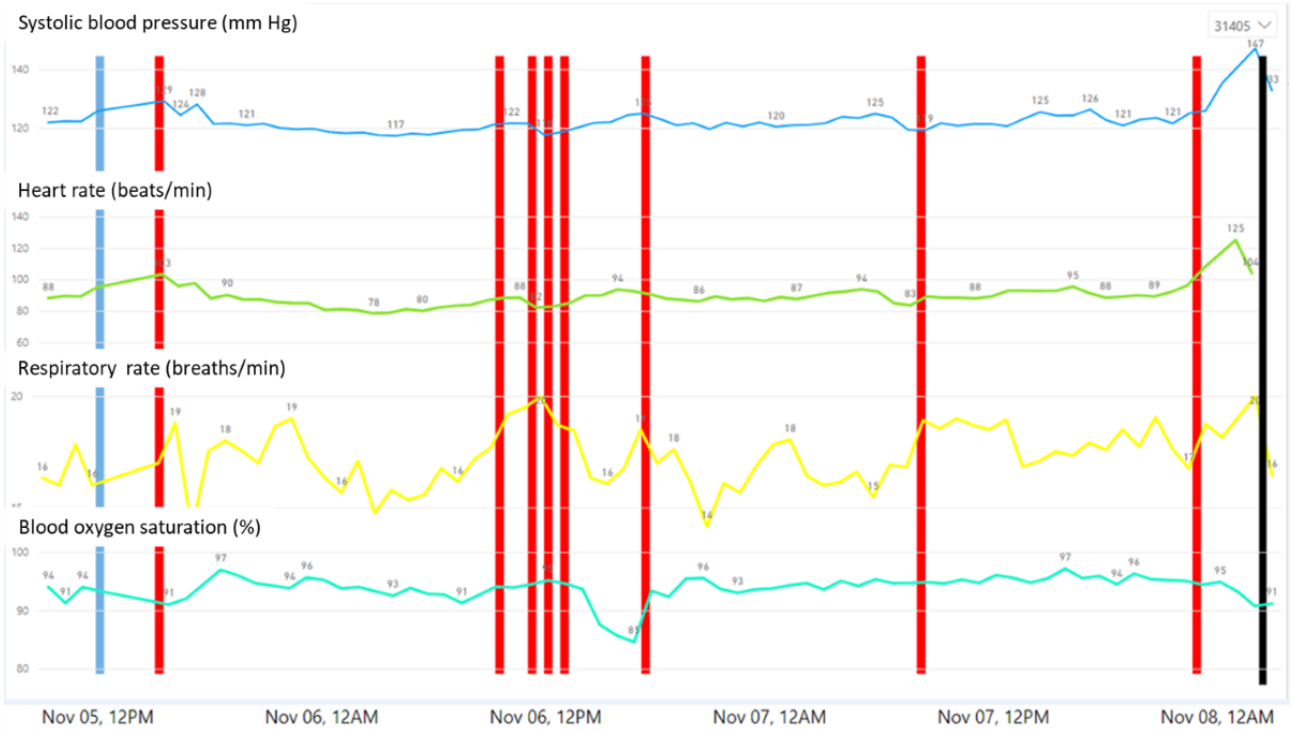


Table 3. Comparison of the sensitivity and specificity of different tools of early detection of deterioration in patients with >200 measurements (N=217).

	Sensitivity, %	Specificity, %
National Early Warning Score	67	22
Airway, Breathing, Circulation, Neurology, and Other score[a]	75	16
The clinical definition of deterioration by local medical staff (“wish list”)	100	0

[a]A locally implemented version of the Airway, Breathing, Circulation, Disability, Exposure criteria for identification of patients' deterioration, as described in Multimedia Appendix 3.

Discussion

Principal Findings

In this study, we have assessed an automated frequent RPM platform with several integrated EWS systems. Alerts were provided many hours before patients have clinically deteriorated; however, similar alerts were also provided for patients who did not deteriorate, questioning the suitability of these EWS systems when frequent monitoring is available. EWS systems are used by health care providers to identify the early signs of clinical deterioration and initiate prompt intervention and management, including nursing staff attention, notifying the clinicians, or activating a rapid response team [27]. A numeric value is assigned to several physiologic parameters, and a composite score is derived and used to identify a patient at risk of deterioration. Most are based on an aggregate weighted system in which the elements are assigned different points for the degree of physiological abnormalities, such as those presented in Multimedia Appendices 2-4. Previous observational studies have suggested that patients often show signs of clinical deterioration up to 24 hours before a serious medical event necessitating an intervention [28]. Delays in care or insufficient treatment of patients on general hospital wards may result in increased admissions to the ICU, cardiac arrest, increased length of hospital stay, or death [28]. The purpose of the EWS scores is to ensure timely and appropriate management of deteriorating patients in general hospital wards. Moreover, for the complex patient population admitted in the general ward and the medical staff treating them, an early warning could be the difference between prevention and late response to decompensation. We also show that when using current early warning systems in a frequent measurement mode, the sensitivity is high, yet the specificity is low, potentially leading to provider fatigue in real-world settings. This is further emphasized when using the "wish list" definitions provided by the medical staff, which has led to an early warning in all 217 patients included in the final analysis, including the 193 patients who had no actual clinical deterioration. This clearly shows that the "wish list" criteria cannot be used for early warning. Previous studies have also shown the relatively high sensitivity of EWS systems; yet, among all patients, the specificity was low [29]. Moreover, in many cases, they provide too many alerts leading to alert fatigue [30].

These results might lead to claims suggesting that clinical judgment is more effective than any EWS system, highlighting the importance of holistic patient care and good clinical judgment. However, it seems that by further improving these EWS systems, sensitivity could be kept high, while specificity would be higher. This was not achieved yet, but preliminary data from various studies implementing big data analysis of multiple physiologic parameters collected automatically and frequently already show promise in early detection of clinically significant changes, and this could eventually result in the desired combination for future EWSs [24,31].

Limitations

Though a limitation of this study is that the health care providers were not using the RPM system in real time and did not react to the warnings it provided, it is expected that in a real-world scenario, once an early warning is provided, the medical staff will intervene and provide the relevant medical support, thus changing the clinical course of the patients from that moment. Importantly, we do not know whether the alerts provided by the different early warning tools resolved spontaneously or whether patients received medical treatment coincidentally at the same time, leading to patient improvement and the resolution of the alert.

At a more practical level, we show that using an RPM system with frequent measurements is feasible in the acute care setting within the general ward. Though 410 patients were recruited, continuous monitoring was achieved properly in only 217 patients (more than 200 measurement sessions) owing to mis-attachment of the wearable monitoring devices. We assume that the reason for that was the blinding of the medical staff from access to the real-time data. Though upon attaching the monitoring devices, the research team ensured that the sensors were well attached and transmitted the data properly, from that moment on, there was no real-time and continuous indication on the quality of the signal. We did see substantial improvement with time, showing that a positive learning curve was rapidly reached and that the devices are simple to use. Moreover, in a real-time scenario, where the medical staff will rely on such a wearable monitoring system, they will immediately receive a notification of an improper signal and will reattach the sensor.

In terms of efficiency, once connected, the data were seamlessly and automatically transferred into the data collection repository and, in parallel, could have been presented on the monitoring screens of the department. This part was not available to the medical staff as they were blinded to real-time monitoring data. EWS compliance is often found to be poor for several reasons, including misinterpretation or incorrect calculation of the scores and poor communications [32]. However, this becomes irrelevant when using an automatically generated and transmitted EWS score.

Further development and future studies are needed to provide an advanced EWS tool that would have higher sensitivity and specificity, making it a better-suited tool in real-world settings, focusing on presymptomatic warnings of potential patient deterioration, to be used as a preventive measure and as a medical decision support tool in both the outpatient and in-hospital settings. The combination of frequently collected multiple physiological parameters, an advanced algorithm, and timely alerts could potentially provide medical staff peace of mind, knowing that they are called only when there is an imminent threat of clinical importance. Moreover, improved prediction of deterioration would have vast positive outcomes when considering the low availability of telemetry beds in hospitals.

Another limitation of the study is that we did not have continuous measurements for all patients during the whole monitoring period. Nonetheless, the data set is much larger than what usually is collected within the general wards, and it still provides important insights. This should be further optimized in future studies on this subject.

Conclusions

To conclude, frequent RPM allows for early detection of physiological changes with potential clinical significance. The integration of an EWS system may provide another layer of clinical awareness, serving as an important decision support tool for early medical intervention. Current scoring systems have high sensitivity but low specificity and warrant further development when combined with frequent multiparameter monitoring. The frequency of measurements alone, though providing a better understanding of trajectories of various vital signs, is not enough to provide an improved EWS score, and in practice, this might be translated into a high rate of alarms, complicating its hospital implementation.

Future systems, which would rely on frequent collection and calculation of the EWS score, could provide better sensitivity and specificity and should be better adjusted to provide tailored scores for the prevention of different medical conditions.

Acknowledgments

This study was funded by the Israel Innovation Authority (grant 64695). The funding source had no active role in this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT Patient and Analysis Flow.
[PNG File , 165 KB - formative_v6i6e36066_app1.png]

Multimedia Appendix 2
National Early Warning Score (NEWS).
[DOCX File , 126 KB - formative_v6i6e36066_app2.docx]

Multimedia Appendix 3
Medical Emergency Team activation criteria.
[DOCX File , 13 KB - formative_v6i6e36066_app3.docx]

Multimedia Appendix 4
Deterioration criteria are defined by the clinical teams of the general wards assuming continuous monitoring available.
[DOCX File , 13 KB - formative_v6i6e36066_app4.docx]

References

1. Rocha HAL, Alcântara ACC, Rocha SGMO, Toscano CM. Effectiveness of rapid response teams in reducing intrahospital cardiac arrests and deaths: a systematic review and meta-analysis. Rev Bras Ter Intensiva 2018;30(3):366-375 [FREE Full text] [doi: 10.5935/0103-507X.20180049] [Medline: 30328990]
2. Bunch JL, Groves PS, Perkhounkova Y. Realistic Evaluation of a Rapid Response System: Context, Mechanisms, and Outcomes. West J Nurs Res 2019 Apr;41(4):519-536. [doi: 10.1177/0193945918776310] [Medline: 29768987]
3. Le Lagadec MD, Dwyer T. Scoping review: The use of early warning systems for the identification of in-hospital patients at risk of deterioration. Aust Crit Care 2017 Jul;30(4):211-218. [doi: 10.1016/j.aucc.2016.10.003] [Medline: 27863876]
4. Chong-Yik R, Bennett AL, Milani RV, Morin DP. Cost-Saving Opportunities with Appropriate Utilization of Cardiac Telemetry. Am J Cardiol 2018 Nov 01;122(9):1570-1573. [doi: 10.1016/j.amjcard.2018.07.016] [Medline: 30193738]
5. Green M, Lander H, Snyder A, Hudson P, Churpek M, Edelson D. Comparison of the Between the Flags calling criteria to the MEWS, NEWS and the electronic Cardiac Arrest Risk Triage (eCART) score for the identification of deteriorating ward patients. Resuscitation 2018 Feb;123:86-91 [FREE Full text] [doi: 10.1016/j.resuscitation.2017.10.028] [Medline: 29169912]
6. Pullinger R, Wilson S, Way R, Santos M, Wong D, Clifton D, et al. Implementing an electronic observation and early warning score chart in the emergency department: a feasibility study. Eur J Emerg Med 2017 Dec;24(6):e11-e16. [doi: 10.1097/MEJ.0000000000000371] [Medline: 26891087]
7. Prgomet M, Cardona-Morrell M, Nicholson M, Lake R, Long J, Westbrook J, et al. Vital signs monitoring on general wards: clinical staff perceptions of current practices and the planned introduction of continuous monitoring technology. Int J Qual Health Care 2016 Sep;28(4):515-521. [doi: 10.1093/intqhc/mzw062] [Medline: 27317251]
8. Tung CE, Su D, Turakhia MP, Lansberg MG. Diagnostic Yield of Extended Cardiac Patch Monitoring in Patients with Stroke or TIA. Front Neurol 2014;5:266 [FREE Full text] [doi: 10.3389/fneur.2014.00266] [Medline: 25628595]

9. Lin B, Wong AM, Tseng KC. Community-Based ECG Monitoring System for Patients with Cardiovascular Diseases. J Med Syst 2016 Apr;40(4):80. [doi: 10.1007/s10916-016-0442-4] [Medline: 26802010]
10. Posthuma LM, Downey C, Visscher MJ, Ghazali DA, Joshi M, Ashrafian H, et al. Remote wireless vital signs monitoring on the ward for early detection of deteriorating patients: A case series. Int J Nurs Stud 2020 Apr;104:103515. [doi: 10.1016/j.ijnurstu.2019.103515] [Medline: 32105974]
11. Muralitharan S, Nelson W, Di S, McGillion M, Devereaux PJ, Barr NG, et al. Machine Learning-Based Early Warning Systems for Clinical Deterioration: Systematic Scoping Review. J Med Internet Res 2021 Feb 04;23(2):e25187 [FREE Full text] [doi: 10.2196/25187] [Medline: 33538696]
12. Watkinson PJ, Barber VS, Price JD, Hann A, Tarassenko L, Young JD. A randomised controlled trial of the effect of continuous electronic physiological monitoring on the adverse event rate in high risk medical and surgical patients. Anaesthesia 2006 Nov;61(11):1031-1039 [FREE Full text] [doi: 10.1111/j.1365-2044.2006.04818.x] [Medline: 17042839]
13. Churpek MM, Adhikari R, Edelson DP. The value of vital sign trends for detecting clinical deterioration on the wards. Resuscitation 2016 May;102:1-5 [FREE Full text] [doi: 10.1016/j.resuscitation.2016.02.005] [Medline: 26898412]
14. Taenzer AH, Perreard IM, MacKenzie T, McGrath SP. Characteristics of Desaturation and Respiratory Rate in Postoperative Patients Breathing Room Air Versus Supplemental Oxygen: Are They Different? Anesth Analg 2018 Mar;126(3):826-832. [doi: 10.1213/ANE.0000000000002765] [Medline: 29293179]
15. Brunetti E, Isaia G, Rinaldi G, Brambati T, De Vito D, Ronco G, et al. Comparison of Diagnostic Accuracies of qSOFA, NEWS, and MEWS to Identify Sepsis in Older Inpatients With Suspected Infection. J Am Med Dir Assoc 2022 May;23(5):865-871.e2. [doi: 10.1016/j.jamda.2021.09.005] [Medline: 34619118]
16. Credland N, Dyson J, Johnson MJ. Do early warning track and trigger tools improve patient outcomes? A systematic synthesis without meta-analysis. J Adv Nurs 2021 Mar;77(2):622-634. [doi: 10.1111/jan.14619] [Medline: 33128308]
17. Bilben B, Grandal L, Søvik S. National Early Warning Score (NEWS) as an emergency department predictor of disease severity and 90-day survival in the acutely dyspneic patient - a prospective observational study. Scand J Trauma Resusc Emerg Med 2016 Jun 02;24:80 [FREE Full text] [doi: 10.1186/s13049-016-0273-9] [Medline: 27250249]
18. Alam N, Vegting IL, Houben E, van Berkel B, Vaughan L, Kramer MHH, et al. Exploring the performance of the National Early Warning Score (NEWS) in a European emergency department. Resuscitation 2015 May;90:111-115. [doi: 10.1016/j.resuscitation.2015.02.011] [Medline: 25748878]
19. Ehara J, Hiraoka E, Hsu H, Yamada T, Homma Y, Fujitani S. The effectiveness of a national early warning score as a triage tool for activating a rapid response system in an outpatient setting: A retrospective cohort study. Medicine (Baltimore) 2019 Dec;98(52):e18475 [FREE Full text] [doi: 10.1097/MD.0000000000018475] [Medline: 31876731]
20. Smith D, Bowden T. Using the ABCDE approach to assess the deteriorating patient. Nurs Stand 2017 Nov 29;32(14):51-63. [doi: 10.7748/ns.2017.e11030] [Medline: 29185642]
21. Nachman D, Gepner Y, Goldstein N, Kabakov E, Ishay AB, Littman R, et al. Comparing blood pressure measurements between a photoplethysmography-based and a standard cuff-based manometry device. Sci Rep 2020 Sep 30;10(1):16116 [FREE Full text] [doi: 10.1038/s41598-020-73172-3] [Medline: 32999400]
22. Nachman D, Constantini K, Poris G, Wagnert-Avraham L, Gertz SD, Littman R, et al. Wireless, non-invasive, wearable device for continuous remote monitoring of hemodynamic parameters in a swine model of controlled hemorrhagic shock. Sci Rep 2020 Oct 19;10(1):17684 [FREE Full text] [doi: 10.1038/s41598-020-74686-6] [Medline: 33077774]
23. Atzmon Y, Ben Ishay E, Hallak M, Littman R, Eisenkraft A, Gabbay-Benziv R. Continuous Maternal Hemodynamics Monitoring at Delivery Using a Novel, Noninvasive, Wireless,PPG-Based Sensor. J Clin Med 2020 Dec 22;10(1) [FREE Full text] [doi: 10.3390/jcm10010008] [Medline: 33375211]
24. Eisenkraft A, Maor Y, Constantini K, Goldstein N, Nachman D, Levy R, et al. Continuous Remote Patient Monitoring Shows Early Cardiovascular Changes in COVID-19 Patients. J Clin Med 2021 Sep 17;10(18) [FREE Full text] [doi: 10.3390/jcm10184218] [Medline: 34575328]
25. Bar-On E, Segal G, Regev-Yochay G, Barkai G, Biber A, Irony A, et al. Establishing a COVID-19 treatment centre in Israel at the initial stage of the outbreak: challenges, responses and lessons learned. Emerg Med J 2021 May;38(5):373-378 [FREE Full text] [doi: 10.1136/emermed-2020-209639] [Medline: 33771818]
26. McKinney W. Data Structures for Statistical Computing in Python. 2010 Presented at: 9th Python in Science Conference (SciPy 2010); June 28 - July 3, 2010; Austin, TX URL: https://pandas.pydata.org/ [doi: 10.25080/Majora-92bf1922-00a]
27. Whittington J, White R, Haig KM, Slock M. Using an automated risk assessment report to identify patients at risk for clinical deterioration. Jt Comm J Qual Patient Saf 2007 Sep;33(9):569-574. [doi: 10.1016/s1553-7250(07)33061-4] [Medline: 17915531]
28. McGaughey J, Alderdice F, Fowler R, Kapila A, Mayhew A, Moutray M. Outreach and Early Warning Systems (EWS) for the prevention of intensive care admission and death of critically ill adult patients on general hospital wards. Cochrane Database Syst Rev 2007 Jul 18(3):CD005529. [doi: 10.1002/14651858.CD005529.pub2] [Medline: 17636805]
29. Forster S, Housley G, McKeever TM, Shaw DE. Investigating the discriminative value of Early Warning Scores in patients with respiratory disease using a retrospective cohort analysis of admissions to Nottingham University Hospitals Trust over a 2-year period. BMJ Open 2018 Jul 30;8(7):e020269 [FREE Full text] [doi: 10.1136/bmjopen-2017-020269] [Medline: 30061434]

30. Kennell TI, Cimino JJ. A Potential Answer to the Alert Override Riddle: Using Patient Attributes to Predict False Positive Alerts. AMIA Annu Symp Proc 2019;2019:532-541 [FREE Full text] [Medline: 32308847]
31. Goldstein N, Eisenkraft A, Arguello CJ, Yang GJ, Sand E, Ishay AB, et al. Exploring Early Pre-Symptomatic Detection of Influenza Using Continuous Monitoring of Advanced Physiological Parameters during a Randomized Controlled Trial. J Clin Med 2021 Nov 08;10(21) [FREE Full text] [doi: 10.3390/jcm10215202] [Medline: 34768722]
32. Le Lagadec MD, Dwyer T, Browne M. The efficacy of twelve early warning systems for potential use in regional medical facilities in Queensland, Australia. Aust Crit Care 2020 Jan;33(1):47-53. [doi: 10.1016/j.aucc.2019.03.001] [Medline: 30979578]

Abbreviations

ABCNO: Airway, Breathing, Circulation, Neurology, and Other
AI: artificial intelligence
EWS: early warning score
ICU: intensive care unit
NEWS: National Early Warning Score
RPM: remote patient monitoring

Edited by A Mavragani; submitted 30.12.21; peer-reviewed by M Le Lagedec, G Rees; comments to author 23.02.22; revised version received 13.03.22; accepted 01.05.22; published 09.06.22.

Please cite as:
Itelman E, Shlomai G, Leibowitz A, Weinstein S, Yakir M, Tamir I, Sagiv M, Muhsen A, Perelman M, Kant D, Zilber E, Segal G
Assessing the Usability of a Novel Wearable Remote Patient Monitoring Device for the Early Detection of In-Hospital Patient Deterioration: Observational Study
JMIR Form Res 2022;6(6):e36066
URL: https://formative.jmir.org/2022/6/e36066
doi:10.2196/36066
PMID:35679119

Original Paper

Characterizing the Experience of Tapentadol Nonmedical Use: Mixed Methods Study

Suzanne K Vosburg[1], PhD; Taryn Dailey-Govoni[1], MPH; Jared Beaumont[1], MPH; Stephen F Butler[1], PhD; Jody L Green[1], PhD

Inflexxion, An Uprise Health | IBH Company, Irvine, CA, United States

Corresponding Author:
Jody L Green, PhD
Inflexxion, An Uprise Health | IBH Company
2 Park Plaza, Suite 1200
Irvine, CA, 92614
United States
Phone: 1 617 765 3138
Email: Jody.Green@uprisehealth.com

Abstract

Background: The prevalence of abuse, diversion, and web-based endorsement of tapentadol (extended-release [ER], immediate-release [IR]) has been characterized as low compared with other prescription opioids. Little is known about individual experience with tapentadol nonmedical use (NMU).

Objective: This study aims to pilot web-based survey technologies to investigate the motivation for tapentadol NMU, sources of procurement, routes of administration, tampering methods, doses used, and impressions of tapentadol products (Nucynta and Nucynta ER).

Methods: Recruitment flyers and banner advertisements were placed on the Bluelight website [DragonByte Technologies Ltd] with a link to a web-based survey (Qualtrics) designed to query about individuals' lifetime tapentadol NMU. This web-based survey was followed by an interactive web-based chat (Cryptocat) with respondents who were willing to be contacted. Respondents were queried about sources for obtaining tapentadol, motives for use, routes of administration, tampering methods, drugs used in combination, tablet strengths and dosages, and reasons for continued or discontinued use. Desirability and attractiveness for NMU was rated.

Results: Web-based recruitment successfully attracted difficult-to-find study participants. A total of 78 participants reported that tapentadol was obtained from friends and family (ER 11/30, 37%; IR 18/67, 27%), the internet (ER 11/30, 37%; IR 12/67, 18%) or participants' own prescriptions from a doctor (ER 9/30, 30%; IR 17/67, 25%). It was used nonmedically for pain relief (ER 18/30, 60%; IR 33/67, 49%) and multiple psychotropic effects, including relaxation (ER 13/30, 43%; IR 29/67, 43%), reduction in depression or anxiety (ER 7/30, 23%; IR 30/67, 45%), or getting high (ER 12/30, 40%; IR 33/67, 49%). Tapentadol was primarily swallowed (ER 22/30, 73%; IR 55/67, 82%), although snorting (ER 2/30, 7%; IR 8/67, 12%) and injection (ER 2/30, 7%; IR 5/67, 8%) were also reported. The preferred dose for NMU was 100 mg (both ER and IR). The participants reported tapentadol use with benzodiazepines (ER 12/21, 57%; IR 28/47, 60%). Most participants had discontinued tapentadol NMU at the time of survey completion (ER 22/30, 73%; IR 55/67, 82%). Reasons for discontinued ER NMU included side effects (10/22, 46%) and lack of effective treatment (10/22, 46%). Reasons for discontinued IR NMU included lack of access (26/55, 47%) and better NMU options (IR 21/55, 38%). Few individuals were willing to divulge identifying information about themselves for the interactive chat (8/78, 10%), demonstrating the strength of anonymous, web-based surveys. Interactive chat supported the survey findings. A subgroup of participants (4/78, 5%) reported hallucinogenic side effects with high doses.

Conclusions: Web-based surveys can successfully recruit individuals who report drug NMU and those who are difficult to find. Tapentadol NMU appears to occur primarily for pain relief and for its psychotropic effects. Although it was liked by some, tapentadol did not receive a robust pattern of endorsement for NMU.

(JMIR Form Res 2022;6(6):e16996) doi:10.2196/16996

KEYWORDS

tapentadol; opioid; prescription opioid; nonmedical use; addiction; chronic pain; web-based survey; Bluelight; drug safety

Introduction

Background

The Controlled Substances Act of the United States regulates the use of drugs by placing them into 1 of 5 descending schedules ranging from schedule I (substances with high potential for abuse or dependence and no accepted medical use in the United States) to schedule V (low potential for abuse or dependence and an accepted medical use in the United States). Tapentadol is a centrally acting, atypical analgesic with a novel mechanism of action that is a combination of μ-opioid agonist activity and norepinephrine reuptake inhibition used to treat moderate to severe acute and chronic pain [1-3]. It was placed in schedule II by the Drug Enforcement Agency because it was determined to have a high potential for abuse, an accepted medical use in treatment in the United States, and the possibility of leading to severe psychological or physical dependence [4]. Tapentadol immediate-release ([IR]; Nucynta) was approved by the United States Food and Drug Administration in December 2008 and the extended-release (ER) formulation (Nucynta ER) in August 2011 [5]. The abuse liability of tapentadol has been of interest since its release, in part because of its proposed mechanism of action and in part because of its initial identification as a schedule II opioid.

Thus far, the prevalence of tapentadol abuse and diversion (both ER and IR) has been characterized as low compared with other prescription opioid compounds, particularly when considered at the population level [6-8]. For instance, from the fourth quarter of 2011 to the second quarter of 2016, tapentadol had an event rate of 0.015 for intentional abuse, an event rate of 0.029 for diversion, and an event rate of 0.245 for past 30-day use to get high. Comparator opioid active pharmaceutical ingredients (hydrocodone, hydromorphone, morphine, oxycodone, oxymorphone, and tramadol) were reported as being intentionally abused from 7.41 (oxymorphone) to 84.32 (oxycodone) times the rate of tapentadol intentional abuse, diverted from 23.172 (oxymorphone) to 316.862 (oxycodone) times the rate of tapentadol diversion and used for getting high in the last 30 days from 3.48 (tramadol) to 52.97 (oxycodone) times the rate of tapentadol [9]. Tapentadol is associated with the fewest serious adverse events among comparator prescription opioid active pharmaceutical ingredients, as well as the fewest dosage units and prescriptions dispensed in the United States [10]. It has also been associated with a lower risk of seeking out multiple physicians to provide prescriptions than oxycodone [11-13]. Nevertheless, when adjusted for prescription volume or drug availability, low levels of abuse-related outcomes are consistently present [9,14].

Objectives

It is difficult to make specific inferences about tapentadol abuse or nonmedical use (NMU) because little has been published on user experiences. This is complicated by the comparatively low rates of tapentadol dispensing in the United States [10], low rates of internet posting about tapentadol compared with other prescription opioids [15], and few behavioral pharmacological studies [16]. Therefore, this study sought to address this gap by piloting web-based recruitment for a tapentadol NMU survey.

Methods

Overview

The Tapentadol Use Internet Survey (TUIS) is a web-based survey followed by an interactive chat among interested survey completers (Multimedia Appendix 1). The survey queried the motivation for tapentadol NMU, sources of drug procurement, routes of administration, tampering methods, doses used, and impressions of tapentadol products (Nucynta and Nucynta ER). The goal of the survey was to solicit detailed information from individuals who self-reported lifetime tapentadol NMU and to pilot the use of web-based recruitment for the difficulty of finding research participants.

Definitions

NMU was defined as the use of tapentadol "in a way not prescribed," including any of the following: (1) used if not prescribed to you; (2) used for reasons other than as a treatment for pain; (3) used via an alternate route of administration (eg, snorted, injected, or other routes not intended for the product); (4) used after tampering (eg, crushed); (5) used in combination with alcohol, illicit drugs, or other prescription drugs without doctor approval; or (6) used at a higher dose than prescribed.

TUIS Survey

Bluelight.org [DragonByte Technologies Ltd] was selected from a pool of drug discussion websites to host the TUIS because at the time of this study its culture supported authentic posts and harm reduction, the site reported a range of approximately 7000 to 10,000 active users within any 30-day period [17], posts were in English, the site was considered stable because it had been in existence for over 10 years, and the staff encouraged research collaborations. A growing number of studies have included data collected from Bluelight.org [18-23].

The 36-item TUIS was developed to solicit information describing tapentadol NMU. Survey construction was an iterative process whereby investigators designed questions to elicit information about participants' prescription opioid use history, and, in particular, their experience with and impressions of tapentadol when used nonmedically, the motivation for tapentadol NMU, sources of procurement, routes of administration, tampering methods, doses used, and general impressions of tapentadol products. The survey was reviewed by Bluelight moderators to ensure their research standards were met [24]. After the investigative and Bluelight teams agreed on content, the survey was posted on Bluelight.org.

Postsurvey Interactive Chat

At the completion of the TUIS, participants were asked to consider chatting interactively with a researcher about their tapentadol product NMU. The chat consisted of 14 questions that provided a framework to enable probing for individual details about initial tapentadol exposure, first use of tapentadol, the high associated with tapentadol, the formulation of tapentadol that had been used for NMU, why it had been used for NMU, number of times using, and the dose of tapentadol most often used for NMU.

Participants and Inclusion/Exclusion Criteria

Individuals participating in the TUIS had to be at least 18 years of age, able to read and understand the English language, reside in the United States, and be willing to provide consent to participate in the survey. They also had to have visited Bluelight.org and report lifetime NMU of a tapentadol product.

Individuals participating in the postsurvey interactive chat had to have completed the TUIS. They had to be willing to provide contact information (email address or Bluelight username) for chatting purposes. Participants had to have the ability to use a web-based chat program and to provide consent to participate.

Ethics Approval

The study was approved for conduct by the New England Institutional Review Board (NEIRB 120170005: Internet Survey and Online Chat Interviews Regarding Tapentadol Use).

Procedures

Participants were recruited for TUIS completion from January through May 2017. A recruitment flyer and banner advertisement were placed on the Bluelight.org website with a link to the web-based survey. This link directed individuals to the consent page which described the voluntary nature of the survey, the absence of payment for survey completion, and information about how to complete the survey. Selecting *I agree to participate* on the informed consent page moved the participant to the beginning of the survey, whereas individuals who did not provide consent to participate in the survey were thanked for their time and brought to an end page. Survey completion took between 5 and 20 minutes depending on the detail of responses provided.

Upon completion of the survey, participants were asked if they would like to be considered for a postsurvey interactive chat regarding their use of tapentadol products. Participants were guaranteed anonymity and asked to provide an email address or a Bluelight username, so they could be contacted by a member of the research team to set up the chat. At the conclusion of the approximately 1-hour, semistructured chat, participants were offered the choice of receiving a US $25 Amazon.com gift certificate or donating the same amount to Erowid.org (a nonprofit drug-education website).

Web-Based Data Collection and Data Analysis

TUIS responses were collected using the web-based data collection software Qualtrics (Qualtrics) and stored in a secure database. The survey was designed for sequential completion such that it was necessary to complete particular items of interest before moving forward in the survey (forced choice), and responses to previously answered questions were carried forward when requesting more detail. A Qualtrics technology function that blocked more than one survey per individual was enabled. Survey items were examined descriptively with frequency and percentages for categorical and binary variables, and means, medians, error, and ranges for continuous variables. Data analysis was carried out using SAS (version 7.11; SAS Institute).

The postsurvey interactive chat was conducted using Cryptocat, a free, open source, encrypted web-based chat program. Transcripts from the postsurvey interactive chats were saved on secure servers with access permissions given only to research staff for data analysis. Thematic data analysis based on grounded theory qualitative research methodology was used to analyze the interview data [25-27]. This is an analytic approach whereby the raw data drive thematic development and analysis.

Specifically, reviewers conducted in-depth review of the chat content to discern and characterize emerging themes regarding tapentadol NMU. In this open coding phase, 2 reviewers (TDG and JB) read the first half of transcripts and assigned topic categories or themes. Constant comparative analysis was performed by repeatedly going back and reviewing previously established categories. While reviewing transcripts, reviewers continually evaluated whether a certain thought or response from an interview participant fit into a previously existing category or whether it represented a unique new theme. This round of review established a basic coding structure.

Subsequently, the 2 reviewers (TDG and JB) read the remaining transcripts and assigned topic categories or themes as appropriate. If additional categories were discerned, the previous transcripts were revisited to ensure that all categories were adequately captured. In this axial category round of analysis (axial coding uses the predefined concepts and categories while rereading the text to confirm that they accurately represent interview responses and to determine how they are related), reviewers established how the categories related to one another. Finally, final selective coding established overall hypotheses and explanations regarding reasons for, opinions of, and experiences with tapentadol NMU.

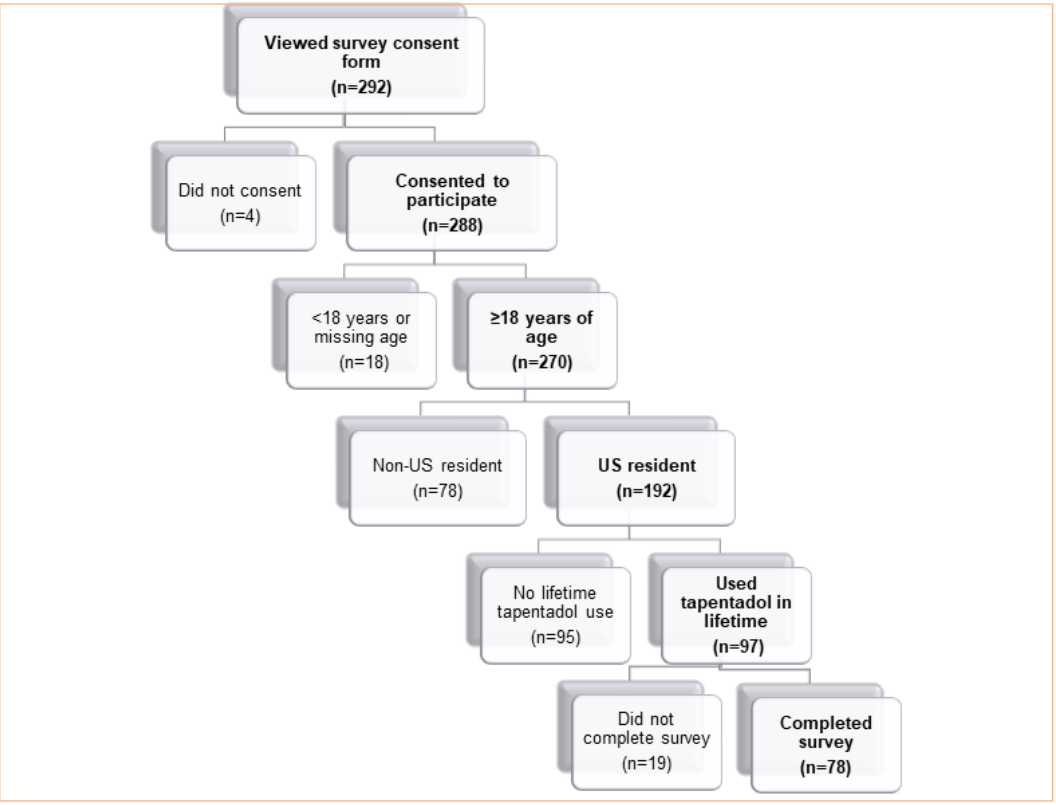
Results

Internet Survey

Participant Characteristics

Figure 1 depicts the disposition of the 78 adults who completed the internet survey from January 2017 to May 2017. Those completing the survey were primarily male (67/78, 86%), White (68/78, 87%), aged between 21 and 54 years (21-34 years 46/78, 59%; 35-54 years 17/78, 22%), with a minimum of some college (61/78, 78%). Most opioids (prescription and illicit) were preferred for NMU; however, 58% (45/78) reported a preference for prescription opioids, opiates, or heroin; 13% (10/78) preferred marijuana and cannabis; 8% (6/78) preferred dissociative drugs; 6% (5/78) (each) preferred psychedelics or prescription stimulants; 4% (3/78) preferred benzodiazepines; 3% (2/78) preferred alcohol; and 1 individual (each) preferred 3,4-methylenedioxymethamphetamine or methamphetamine.

Figure 1. Disposition of survey respondents.



A total of 3 participants reported opioid NMU solely with tapentadol products, whereas the remainder of the sample reported lifetime prescription opioid NMU with other opioids in addition to tapentadol. Most began using prescription opioids nonmedically between the ages of 14 and 18 years (44/78, 57%), followed by those aged 19 to 25 years (16/78, 21%), and their first NMU of prescription opioids was hydrocodone IR (29/78, 37%), oxycodone IR (13/78, 17%), or codeine (11/78, 14%). Most patients (45/78, 58%) still used opioids nonmedically.

Lifetime Tapentadol NMU

According to the inclusion criteria, all participants reported lifetime NMU of tapentadol: 39% (30/78) reported tapentadol ER NMU, and 86% (67/78) reported tapentadol IR NMU. Almost one-fourth of the sample reported NMU with both ER and IR formulations (19/78, 24%), whereas 14% (11/78) reported NMU solely with tapentadol ER, and 62% (48/78) reported NMU solely with tapentadol IR.

Age at First Episode of Tapentadol NMU

Approximately 40% of both tapentadol ER (12/30) and IR (27/67) users reported that they were aged >25 years when they first used tapentadol nonmedically. The remaining ages of the first tapentadol ER NMU were between 14 and 18 (9/30, 30%) years or 19 and 25 (8/30, 27%) years, with an individual aged <10 years. The remaining ages of the first tapentadol IR NMU were between 19 and 25 (25/67, 37%) and 14 and 18 (13/67, 19%) years.

Procurement

Table 1 presents the procurement sources of tapentadol for NMU. Tapentadol was often given to participants by family members, friends, or acquaintances (ER 11/30, 37%; IR 18/67, 27%). It was also obtained using the internet (ER 11/30, 37%; IR 12/67, 18%) or from the participants' own prescriptions from a doctor (ER 9/30, 30%; IR 17/67, 25%). Other sources included being stolen (ER 4/30, 13%), being bought from a dealer (IR 11/67, 16%), or being bought from friends or family (IR 9/67, 13%).

Table 1. Sources of tapentadol for nonmedical use organized by formulation.

Source	Tapentadol extended-release (N=30), n (%)	Tapentadol immediate-release (N=67), n (%)
Given to me by a family member, friend, or acquaintance	11 (37)	18 (27)
Internet sources	11 (37)	12 (18)
Own prescription from one doctor	9 (30)	17 (25)
Stolen	4 (13)	6 (9)
Bought from a dealer (someone known to sell drugs)	2 (7)	11 (16)
Bought from a family member, friend, or acquaintance	2 (7)	9 (13)
Own prescription from multiple doctors	1 (3)	1 (2)
Other	1 (3)	3 (5)

Motives for Use

Table 2 presents the motives for using tapentadol NMU. Both formulations were mostly used nonmedically for pain relief (ER 18/30, 60%; IR 33/67, 49%) or to feel high, buzzed, or stoned (IR 33/67, 49%). The remaining main motives for NMU included relaxation (ER 13/30, 43%; IR 29/67, 43%) and feeling less depressed or anxious (ER 7/30, 23%; IR 30/67, 45%). Other motives included treatment or prevention of withdrawal symptoms (ER 6/30, 20%; IR 18/67, 27%), treatment of emotional pain (ER 5/30, 17%; IR 26/67, 39%), or feeling more outgoing (ER 6/30, 20%; IR 17/67, 25%) or energetic (ER 6/30, 20%; IR 16/67, 24%).

Table 2. Motives for tapentadol nonmedical use organized by formulation.

Motive	Tapentadol extended-release (N=30), n (%)	Tapentadol immediate-release (N=67), n (%)
To provide better pain relief	18 (60)	33 (49)
To relax	13 (43)	29 (43)
To feel high or buzzed or stoned	12 (40)	33 (49)
To feel less depressed or anxious	7 (23)	30 (45)
To feel more outgoing	6 (20)	17 (25)
To feel more energetic	6 (20)	16 (24)
To treat or prevent withdrawal symptoms	6 (20)	18 (27)
To treat emotional pain	5 (17)	26 (39)
To enhance the recreational effects of other drugs or substances	3 (10)	7 (10)
To reduce stress	3 (10)	26 (39)
To ease the comedown from other drugs or substances	2 (7)	7 (10)
To experience psychedelic effects (eg, hallucinations)	1 (3)	3 (5)
Other reason[a]	1 (3)	7 (10)

[a]Tapentadol extended-release other reason: curiosity n=1; tapentadol immediate-release other reasons: curiosity n=5, sexual reasons n=1, chronic Fatigue n=1.

Routes of Administration

Table 3 summarizes the routes of administration of NMU tapentadol. Tapentadol was predominantly administered orally, and all patients who swallowed pills reported using alternate oral routes of administration. The participants reported swallowing pills (ER 22/30, 73%; IR 55/67, 82%) or chewing (ER 6/30, 20%; IR 17/67, 25%). Both formulations were ingested using the parachute technique, which refers to wrapping crushed pills in any type of digestible paper (toilet paper or tissue paper) to avoid a bitter taste (ER 6/30, 20%; IR 7/67, 10%). Nonoral routes of administration were used less frequently. Participants reported their preferred route of administration as swallowing whole for each formulation (ER 17/30, 57%; IR 44/67, 66%), followed by chewing (ER 3/30, 10%; IR 7/67, 11%).

Table 3. Routes of administration ever used and preferred route of administration for tapentadol organized by formulation.

Route	Tapentadol extended-release (N=30)		Tapentadol immediate-release (N=67)	
	Used, n (%)	Preferred, n (%)	Used, n (%)	Preferred, n (%)
Swallow whole	22 (73)	17 (57)	55 (82)	44 (66)
Chew	6 (20)	3 (10)	17 (25)	7 (11)
Parachute	6 (20)	2 (7)	7 (10)	4 (6)
Drink in solution	3 (10)	2 (7)	4 (6)	1 (2)
Sublingual	2 (7)	1 (3)	6 (9)	3 (5)
Inject	2 (7)	0 (0)	5 (8)	4 (6)
Snort	2 (7)	1 (3)	8 (12)	2 (3)
Other oral	2 (7)	2 (7)	0 (0)	0 (0)
Rectal	2 (7)	2 (7)	5 (8)	2 (3)
Buccal	1 (3)	0 (0)	1 (2)	0 (0)
Smoke	1 (3)	0 (0)	0 (0)	0 (0)

Tampering

Table 4 summarizes the tapentadol NMU tampering methods. A subset reported no medication tampering before NMU (ER 9/30, 30%; IR 33/67, 49%). The tampering strategies most frequently used with the ER formulation were breaking into smaller pieces (12/30, 40%), crushing or grinding or shaving (8/30, 27%), chewing (7/30, 23%), or dissolving or soaking (6/30, 20%). Similarly, the most frequently used strategies for IR formulation tampering included crushing or grinding or shaving (16/67, 24%), chewing (15/67, 22%), breaking into smaller pieces (12/67, 18%), and soaking or dissolving (6/67, 9%).

Table 4. Tampering methods for nonmedical use (NMU) organized by formulation.

Tampering method	Tapentadol extended-release (N=30), n (%)	Tapentadol immediate-release (N=67), n (%)
No tampering before NMU	9 (30)	33 (49)
Break into smaller pieces	12 (40)	12 (18)
Crush or grind or shave	8 (27)	16 (24)
Chew	7 (23)	15 (22)
Soak or dissolve	6 (20)	6 (9)
Filter dissolved product in liquid using wheel filter, micron filter, or other type of filter	3 (10)	3 (5)
Heat	1 (3)	3 (5)
Cool or freeze	1 (3)	0 (0)
Filter dissolved product in liquid using coffee ball, coffee filter, or other material	0 (0)	3 (5)
Other[a]	0 (0)	1 (2)

[a]Under tapentadol immediate-release, other refers to Dremel tool.

Drug Combinations for NMU

The use of tapentadol in combination with other drugs was reported by 70% (21/30) of those who used tapentadol ER for NMU and 70% (47/67) of those who used tapentadol IR for NMU. Table 5 shows that the patterns of drug combinations were similar across both formulations. Benzodiazepines were the most frequently reported drugs used in both formulations (ER 12/21; 57%; IR 28/47, 60%). Alcohol was used at the same rate with ER formulations as benzodiazepines were, but it was used less with IR formulations (IR 18/47, 38%). Other drugs used in combination with tapentadol products were prescription opioids (ER 11/21, 52%; IR 20/47, 43%), marijuana or cannabis (ER 9/21, 43%; IR 17/47, 36%), or prescription stimulants (ER 5/21, 24%; IR 9/47, 19%). Heroin (ER 4/21, 19%; IR 5/47, 11%) and cocaine (ER 2/21, 10%; IR 5/47, 11%) were also taken along with tapentadol.

Table 5. Drugs used in combination with tapentadol for nonmedical use[a].

	Tapentadol extended-release (N=21), n (%)	Tapentadol immediate-release (N=47), n (%)
Benzodiazepines	12 (57)	28 (60)
Alcohol	12 (57)	18 (38)
Rx opioids	11 (52)	20 (43)
Marijuana or cannabis	9 (43)	17 (36)
Rx stimulants	5 (24)	9 (19)
Heroin	4 (19)	5 (11)
Cocaine	2 (10)	5 (11)
Hallucinogens or psychedelics	2 (10)	3 (6)
3,4-methylenedioxymethamphetamine or empathenogenic drugs	1 (5)	2 (4)
Methamphetamine	1 (5)	3 (6)
Antidepressants	1 (5)	5 (11)
Other	0 (0)	4 (9)
Dissociative drugs	0 (0)	0 (0)
Inhalants	0 (0)	2 (4)
Bath salts or 3,4-methylenedioxypyrovalerone	0 (0)	0 (0)

[a]Percentages were calculated with those who reported using drugs in combination with tapentadol.

Strength of Tapentadol Tablets Used for NMU and Highest Milligram Amount in One Use Session

Among participants who reported tapentadol ER NMU 57% (17/30) used 100 mg tablets followed by 50 mg and 200 mg strength tablet (33%, 10/30, each). Among participants who reported tapentadol IR NMU, 54% (36/67) used the 100 mg tablet followed by the 75 mg tablet (34%, 23/67) and the 50 mg tablet (31%, 21/67). The pattern of highest milligram amount used in a session was also similar between ER and IR formulations: approximately one-third of ER and IR users reported using between 100 and 200 mg during 1 session (ER 10/30, 33%; IR 24/67, 36%), followed by 251 mg to 500 mg (ER 8/30, 27%; IR 16/67, 24%).

End of Use

Most participants with tapentadol ER (22/30, 73%) or IR (55/67, 82%) lifetime NMU no longer used the products at the time of survey completion. Tables 6 and 7 reveal that the primary reasons for discontinuing ER NMU included negative side effects (ER 10/22, 46%), an ineffective high (ER 10/22, 46%), not liking the way the drug felt (ER 8/22, 36%), or ineffective pain relief (ER 8/22, 36%). The primary reasons for discontinued IR use were lack of access (IR 26/55, 47%), availability of better options (IR 21/55, 38%), or an ineffective high (IR 18/55, 33%).

Table 6. Reasons for discontinued nonmedical use of tapentadol extended-release.

	Tapentadol extended-release (N=22), n (%)	Tapentadol immediate-release (N=55), n (%)
Have experienced negative side effects	10 (46)	15 (27)
Drug does not provide an effective high	10 (46)	18 (33)
Do not enjoy the way the drug makes me feel (Buzz or Nod)	8 (36)	15 (27)
Drug not effective at pain relief	8 (36)	15 (27)
Better options are available	7 (32)	21 (38)
Tapentadol is too expensive	6 (27)	11 (20)
Do not have access to the drug	6 (27)	26 (47)
Worried about negative side effects	3 (14)	11 (20)
Do not enjoy the psychedelic effects of the drug	2 (9)	4 (7)
Difficult to manipulate product to use via my preferred route of administration	2 (9)	1 (2)
More stigma about this drug than other drugs	1 (5)	3 (6)

Table 7. Reasons for continued nonmedical use of tapentadol extended-release and immediate-release.

	Tapentadol extended-release (N=8), n (%)	Tapentadol immediate-release (N=12), n (%)
Drug is effective at pain relief	6 (75)	8 (67)
Enjoy how the drug makes me feel	3 (38)	4 (33)
I have access to the drug	2 (25)	7 (58)
I do not have better options	2 (25)	4 (33)
The drug provides an effective high	2 (25)	3 (25)
Drug is inexpensive	1 (13)	4 (33)
Enjoy psychedelic effects	1 (13)	1 (8)
Not worried about negative effects	1 (13)	2 (17)
Easy to manipulate for my preferred route of administration	1 (13)	1 (8)
Less stigma of this drug	1 (13)	1 (8)
I have not experienced any negative side effects	0 (0)	3 (25)

A small number of participants reported continued NMU of tapentadol ER (8/30, 27%) and IR (12/67, 18%; Table 7). The primary reason for continued NMU in both ER (6/8, 75%) and IR (8/12, 67%) was effective pain relief. The remaining primary reasons were enjoyment of the way the drug felt (ER 3/8, 38%; IR 4/12, 33%) or having access to the drug (ER 2/8, 25%; IR 7/12, 58%).

Desirability Ratings of Tapentadol

Survey participants were asked to rate the desirability of opioids they had used nonmedically on a scale ranging from 1 to 100, with 100 representing the best drug imaginable for NMU and 1 representing the worst drug imaginable for NMU. The median values are presented because the number of ratings ranged from 10 (hydromorphone ER) to 67 (tapentadol IR). Table 8 summarizes these findings and illustrates that, compared with other prescription opioid compounds, the median desirability ratings for tapentadol were relatively low (ER=37; IR=41). The highest ratings were for oxymorphone IR (96) and the lowest ratings were for tramadol ER (19).

Table 8. Median desirability ratings of opioids used nonmedically.

API[a]	Desirability Rating (Median)	Number of rating and percentage of sample (N=78), n (%)
Oxymorphone IR[b]	96	21 (27)
Oxycodone IR noncombination	90	47 (60)
Oxymorphone ER[c]	86	17 (22)
Hydromorphone IR	85	37 (47)
Morphine IR	84	33 (42)
Oxycodone ER	80	44 (56)
Other	80	16 (21)
Oxycodone IR combination	78	56 (72)
Hydromorphone ER	72	10 (13)
Hydrocodone IR	68	62 (79)
Fentanyl	62	41 (53)
Methadone	61	34 (44)
Morphine ER	60	33 (42)
Hydrocodone ER	53	12 (15)
Meperidine	43	13 (17)
Tapentadol IR	41	67 (86)
Tapentadol ER	37	30 (38)
Codeine	35	49 (63)
Buprenorphine	33	32 (41)
Tramadol IR	24	53 (68)
Tramadol ER	19	22 (28)

[a]API: active pharmaceutical ingredient.

[b]IR: immediate-release.

[c]ER: extended-release.

Interactive Web-Based Chat

A total of 8 survey participants (10% of the survey sample) agreed to participate in a follow-up semistructured interactive web-based chat. The demographic profile was similar to that of the full sample: 7 (88%) were male, 6 (75%) were White, 6 (88%) were aged <27 years, and 7 (88%) attended a minimum of some college. All participants reported tapentadol IR NMU and half (4/8, 50%) reported tapentadol ER NMU. A total of 6 (75%) participants reported prescription opioids or heroin as their preferred drug for NMU. The remaining 2 (25%) participants reported marijuana and psychedelics (n=1) and dissociative drugs (n=1) as drugs of choice.

Table 9 summarizes the themes and supporting statements derived from participants' descriptive responses to questions regarding their tapentadol NMU experience. Sentiments and side effects were coded positive if they were favorable or neutral if participants indicated that they neither liked nor disliked the experience or liked one aspect but disliked another. Negative experiences were coded as such if the experience was clearly disliked and typically included extreme or no effects at higher doses.

Of the 8 participants, 3 (38%) reported favorable or positive sentiment, 3 (38%) reported neutral sentiment, and 2 (25%) reported negative sentiment. The side effects of tapentadol were described solely in neutral to negative terms. Reported tampering efforts reflected the composition of the ER formulation (Nucynta ER is formulated with inactive ingredients that make it difficult to crush, although it is not recognized by the Food and Drug Administration as having abuse-deterrent properties): "Nucynta [ER] was like a solid chunk of plastic." They also reflected the rationale for the oral use of both ER ("I could barely change its shape in the slightest [and] ended up swallowing it whole") and IR formulations ("If I don't need to beat a time-release formulation, I typically just swallow 95% of the time"). Comments also reflected the undesirability of the IR formulation for alternate routes of administration: "We actually all tried to snort it… [but] it burned and hurt superbly" (Table 8). Finally, tapentadol was referred to as a fairly obscure find in traditional diversion settings. An individual reported that when trying to sell it or even share it, people "wouldn't have anything to do with it" because they did not recognize the product.

Table 9. Emergent themes from semistructured interview.

Theme	Quotes
Multiple and varying opinions were expressed about the recreational qualities of tapentadol	
Positive opinions	• ER was great after I threw up; both ER and IR felt pretty much the same as any other opioid: fuzzy head and body, warm limbs, euphoric... • ER [floaty and euphoric; speedy and uplifting], was not as abusable [sic] as IR since it took a long time to peak and lasted so long. IR had intense sedation and euphoria, but [I was] awake during the experience. Strongest opioid feeling I ever felt/Perfect opioid/fell in love with it • Liked it but high was unusual—almost stimulating but not really… like maybe half a shot of espresso? ... Hard to say… Very little nausea and histamine reaction, ie, not a lot of itching was a plus, the pills themselves were pretty strong individually and were small [did not have to swallow a bunch of big-ass Percocets, for example]... there was not anything I disliked about it specifically over other opiates…
Neutral opinions	• I cannot say I have any likes or dislikes [about Nucynta ER]: No high whatsoever, slightly more effective than NSAIDs, slightly less effective than even codeine. IR—no high, liked that it helped with withdrawal relief [mild], pain relief [mild]; negative effects are preventive • Cannot say it was really pleasant, it's very overwhelming… Attracted to it for psychedelic potential and novelty of experience; Lower doses helped tremendously with anxiety; Was unique • As enjoyable as any other opioid at the equianalgesic dose. Experience was underwhelming, minor mitigation of withdrawal symptoms, similar to how we use codeine
Negative opinions	• Underwhelmed, but kept at it: the higher doses were more rewarding, but the more I used it, the more bad effects I had • Dissimilar to other opioids, Costs a lot, no effects, worthless for pain
Tapentadol had side effects described as neutral to negative (5 out of 8 participants)	• Auditory hallucinations were neither desirable or undesirable • At high dose became extremely visual and extremely disorientating, extremely nauseating. Dislike nausea… • Led to hallucinations, no fun • Negative effects at high IR dosage would prevent me from using again. Very mild mu opiate with overpowering negative effects [IR felt like kappa agonism: slight dysphoria, slight dissociation] • Must stress that when used by itself at too high [a dose] the side effects can be really bad. I spent a couple hours with weird zaps in my head, cloudy thoughts, messed up speech—it was bad and those close to me were worried. It's anti-abusive nature actually opens up some really crazy effects that were pretty much my worst response to any pain medication
Tapentadol was mostly used orally; tampering efforts were limited to simple methods	• Once I learned about soaking it in a carbonated drink for a while, I took the ERs like that • I couldn't crush it [...] the Nucynta [ER] was like a solid chunk of plastic. I could barely change its shape in the slightest. I ended up swallowing it whole • Pills swallowed whole/Just took it orally/When it came to pills, if I do not need to beat a time-release formulation, I typically just swallow 95% of the time • We actually all tried to snort [insufflate] it [Nucynta IR] through our noses, but we quickly realized that was not an option because it burned and hurt superbly. So, we all took it orally […] I swallowed it whole. It is such a bitter compound that it isn't worth chewing up in my opinion • I would cut the [Nucynta ER] pills into quarters with a razor blade and swallow the pieces… it took a little effort
Tapentadol is not well known in traditional diversion settings	• It was available [through an] acquaintance who was a pain patient [and] had just had it prescribed. It did not come from a "dealer" per se and I have never seen nor heard of it coming from such a source • It's nowhere on the street, 99% of people have not heard of it and the few times… I tried to share/sell any people would not have anything to do with it because they did not know what it was

Discussion

Principal Findings

To date, studies on the abuse liability of tapentadol have focused on aggregate outcome measures [6,7,9,14,15,28]. The present survey sought to address a gap in the tapentadol literature by soliciting direct feedback from individuals with tapentadol NMU experience and characterize the associated motivations, behaviors, and consequences of NMU. To do so, web-based recruitment and survey technology were piloted, and it was found that they were an effective method to recruit a difficult-to-find research sample.

Similar to other prescription opioids [29], the main source of tapentadol was friends, family, or acquaintances. Tapentadol was also obtained on the web and through other sources of diversion, such as being stolen, drug dealers, or purchased directly from friends or family. Although many sources have not reported significant levels of diversion [7,9,14,15], these data reveal a type of tapentadol diversion that is occurring, although perhaps at low levels.

It was hypothesized that individuals might use tapentadol for reasons other than analgesia, such as the rumored psychedelic effects [5,30,31] (Tables 6-8). However, the primary reason for tapentadol NMU and ongoing tapentadol NMU across both

formulations was better pain relief, followed by psychotropic effects, including relaxation, reduction in depression or anxiety, or getting high. Approximately 25% to 30% of participants reported that they misused their own tapentadol prescription, revealing NMU among some patients with pain. Pain has been found to contribute to the risk of developing prescription opioid use disorders over time [32], and it is possible that the present data capture aspects of this relationship.

Benzodiazepines are most often used in combination with both formulations of tapentadol. Benzodiazepines can pose a life-threatening risk when used concomitantly with opioids because of the increased risk of respiratory depression and overdose [33,34]. Even so, they are prescribed at varying rates to patients undergoing opioid maintenance therapy. Individuals seeking opiate detoxification also report using benzodiazepines to manage anxiety, help with sleep, decrease opioid withdrawal, enhance the recreational effects of other drugs or substances, or get high [35]. The rationale for the concomitant use of benzodiazepine and tapentadol was not discussed in this study.

Individuals who used tapentadol at high doses (≥200 mg) reported hallucinations. Some interview participants did not identify these effects as positive or negative, but others reported them as strong deterrents to future tapentadol NMU. Hallucinations are described as part of a serotonin syndrome that can occur when taking serotonin and norepinephrine reuptake inhibitor products such as tapentadol [5] and in combination with other serotonergic drugs [36]. To date, the literature is inconclusive as to whether tapentadol alone has resulted in a true serotonin syndrome experience [30,31].

Desirability ratings were lower for tapentadol than for the other opioid compounds. Ratings of desirability were similar between tapentadol ER and IR, with tapentadol ER being less attractive for NMU than IR. In addition to supporting recent findings that IR formulations are more desirable for NMU than ER formulations [37], these data also suggest a lack of desirability for the entire tapentadol molecule, not just for one formulation. In further support of this inference, most participants (61/78; 78%) stopped using either formulation at the time of the survey. These findings were similar to those reported in the study by McNaughton et al [15]. Regardless of how individuals reported using tapentadol, most participants (ER 22/30, 73%; IR 55/67, 82%) did not indicate that they enjoyed it.

Limitations

Limitations of this study include the use of self-reported data from a self-selected, US-based, convenience sample responding to a pilot internet survey, which may not be a fully representative sample of nonmedical users of tapentadol. The selection of individuals who reported lifetime tapentadol use meant that some did not report their current experiences. The modest number of participants may be because of the market share of tapentadol, but it may also be because of the nondesirability of tapentadol for NMU. This may be a topic for future research [8]. Extension of the recruitment period longer than 5 months might result in a larger sample size, if the reason for the modest sample is the lack of tapentadol market penetration. Notably, few participants volunteered to participate in the follow-up survey. Although it is possible that this activity may not have been of interest, it may also have been because of the requirement to provide an email address or a Bluelight username (which suggests indirect support for anonymous web-based surveys). However, it also suggests an opportunity for another technological development in which survey completers could remain anonymous yet respond to specific follow-up questions. Much of what was reported is similar to other findings in the literature documented herein, lending face value to this report. New directions for future surveys may include patterns of tapentadol NMU, such as frequency of use, redosing, and the degree to which larger doses are used to obtain the same effect. Finally, participants were also aware that the focus of the study was tapentadol and may have felt compelled to over- or underreport its use. Even so, great care was taken to ensure the quality of data collection and analysis.

Conclusions

In conclusion, these preliminary data reveal potential avenues for further exploration of NMU tapentadol. The use of web-based survey technology for survey recruitment of a difficult-to-find sample and a follow-up interactive chat may be another useful technology for postmarketing surveillance studies. The primary motive for continuing tapentadol NMU was pain relief. Tapentadol ER (12/21, 57%) or IR (28/47, 60%) use together with benzodiazepines were reported . There is also some evidence of diversion. At high doses, psychotropic effects have been reported. Most NMU of tapentadol occurred via oral routes of administration. Similar to other studies, although it was liked by some, tapentadol did not receive a robust pattern of endorsement for NMU.

Conflicts of Interest

Authors TDG, JB and JLG are employees of and SFB is a consultant to Inflexxion, an Uprise |IBH Company. At the time this paper was written, author SKV was an independent scientific writer and consultant contracting with multiple companies. She has since joined Inflexxion. Inflexxion contracts with the Federal Drug Administration (FDA) and multiple companies with interests in some of the products included in the compounds evaluated for this article. Funding for this research was provided initially by Depomed, Inc. and later continued by Collegium Pharmaceutical, Inc. Although the sponsor was involved in reviewing the content of this article, all data collection, analysis, and ultimate data interpretation were made by the authors without sponsor influence

Multimedia Appendix 1
Tapentadol use internet survey.
[PDF File (Adobe PDF File), 390 KB - formative_v6i6e16996_app1.pdf]

References

1. Tzschentke TM, Christoph T, Kögel BY. The mu-opioid receptor agonist/noradrenaline reuptake inhibition (MOR-NRI) concept in analgesia: the case of tapentadol. CNS Drugs 2014 Apr;28(4):319-329. [doi: 10.1007/s40263-014-0151-9] [Medline: 24578192]
2. Tzschentke TM, Jahnel U, Kogel B, Christoph T, Englberger W, De Vry J, et al. Tapentadol hydrochloride: a next-generation, centrally acting analgesic with two mechanisms of action in a single molecule. Drugs Today (Barc) 2009 Jul;45(7):483-496. [doi: 10.1358/dot.2009.45.7.1395291] [Medline: 19834626]
3. Raffa RB, Buschmann H, Christoph T, Eichenbaum G, Englberger W, Flores CM, et al. Mechanistic and functional differentiation of tapentadol and tramadol. Expert Opin Pharmacother 2012 Jul;13(10):1437-1449. [doi: 10.1517/14656566.2012.696097] [Medline: 22698264]
4. Drug Enforcement Administration, Department of Justice. Schedules of controlled substances: placement of tapentadol into schedule II. Final rule. Fed Regist 2009 May 21;74(97):23790-23793 [FREE Full text] [Medline: 19507329]
5. Stollenwerk A, Sohns M, Heisig F, Elling C, von Zabern D. Review of post-marketing safety data on Tapentadol, a centrally acting analgesic. Adv Ther 2018 Jan;35(1):12-30 [FREE Full text] [doi: 10.1007/s12325-017-0654-0] [Medline: 29270779]
6. Dart RC, Cicero TJ, Surratt HL, Rosenblum A, Bartelson BB, Adams EH. Assessment of the abuse of tapentadol immediate release: the first 24 months. J Opioid Manag 2012;8(6):395-402. [doi: 10.5055/jom.2012.0139] [Medline: 23264317]
7. Dart RC, Surratt HL, Le Lait MC, Stivers Y, Bebarta VS, Freifeld CC, et al. Diversion and illicit sale of extended release Tapentadol in the United States. Pain Med 2016 Aug;17(8):1490-1496 [FREE Full text] [doi: 10.1093/pm/pnv032] [Medline: 26814267]
8. Vosburg SK, Beaumont J, Dailey-Govoni ST, Butler SF, Green JL. Evaluation of abuse and route of administration of extended-release Tapentadol among treatment-seeking individuals, as captured by the Addiction Severity Index-Multimedia Version (ASI-MV). Pain Med 2020 Sep 01;21(9):1891-1901 [FREE Full text] [doi: 10.1093/pm/pnz250] [Medline: 31617931]
9. Vosburg SK, Severtson SG, Dart RC, Cicero TJ, Kurtz SP, Parrino MW, et al. Assessment of Tapentadol API abuse liability with the researched abuse, diversion and addiction-related surveillance system. J Pain 2018 Apr;19(4):439-453. [doi: 10.1016/j.jpain.2017.11.007] [Medline: 29224919]
10. Murphy DL, Lebin JA, Severtson SG, Olsen HA, Dasgupta N, Dart RC. Comparative rates of mortality and serious adverse effects among commonly prescribed opioid analgesics. Drug Saf 2018 Aug;41(8):787-795. [doi: 10.1007/s40264-018-0660-4] [Medline: 29582394]
11. Cepeda MS, Fife D, Kihm MA, Mastrogiovanni G, Yuan Y. Comparison of the risks of shopping behavior and opioid abuse between Tapentadol and Oxycodone and association of shopping behavior and opioid abuse. Clin J Pain 2014 Dec;30(12):1051-1056 [FREE Full text] [doi: 10.1097/AJP.0000000000000067] [Medline: 24370606]
12. Cepeda MS, Fife D, Ma Q, Ryan PB. Comparison of the risks of opioid abuse or dependence between Tapentadol and Oxycodone: results from a cohort study. J Pain 2013 Oct;14(10):1227-1241 [FREE Full text] [doi: 10.1016/j.jpain.2013.05.010] [Medline: 23850177]
13. Cepeda MS, Fife D, Vo L, Mastrogiovanni G, Yuan Y. Comparison of opioid doctor shopping for Tapentadol and Oxycodone: a cohort study. J Pain 2013 Feb;14(2):158-164 [FREE Full text] [doi: 10.1016/j.jpain.2012.10.012] [Medline: 23253635]
14. Butler SF, McNaughton EC, Black RA. Tapentadol abuse potential: a postmarketing evaluation using a sample of individuals evaluated for substance abuse treatment. Pain Med 2015 Jan;16(1):119-130. [doi: 10.1111/pme.12524] [Medline: 25243972]
15. McNaughton EC, Black RA, Weber SE, Butler SF. Assessing abuse potential of new analgesic medications following market release: an evaluation of Internet discussion of Tapentadol abuse. Pain Med 2015 Jan;16(1):131-140. [doi: 10.1111/pme.12547] [Medline: 25244069]
16. Stoops WW, Glaser PE, Rush CR. Miotic and subject-rated effects of therapeutic doses of Tapentadol, Tramadol, and Hydromorphone in occasional opioid users. Psychopharmacology (Berl) 2013 Jul;228(2):255-262 [FREE Full text] [doi: 10.1007/s00213-013-3031-y] [Medline: 23430163]
17. The Bluelight Research Portal. Bluelight. 2019 May 16. URL: https://www.bluelight.org/xf/pages/Research_Portal/ [accessed 2022-05-06]
18. Chiauzzi E, Dasmahapatra P, Lobo K, Barratt MJ. Participatory research with an online drug forum: a survey of user characteristics, information sharing, and harm reduction views. Subst Use Misuse 2013 Jun;48(8):661-670. [doi: 10.3109/10826084.2013.800117] [Medline: 23750771]
19. Deligianni E, Corkery JM, Schifano F, Lione LA. An international survey on the awareness, use, preference, and health perception of novel psychoactive substances (NPS). Hum Psychopharmacol 2017 May;32(3):e2581. [doi: 10.1002/hup.2581] [Medline: 28485125]
20. Deligianni E, Daniel OJ, Corkery JM, Schifano F, Lione LA. Impact of the UK Psychoactive Substances Act on awareness, use, experiences and knowledge of potential associated health risks of novel psychoactive substances. Br J Clin Pharmacol 2020 Mar;86(3):505-516 [FREE Full text] [doi: 10.1111/bcp.14123] [Medline: 31495968]
21. Lamy FR, Daniulaityte R, Nahhas RW, Barratt MJ, Smith AG, Sheth A, et al. Increases in synthetic cannabinoids-related harms: results from a longitudinal web-based content analysis. Int J Drug Policy 2017 Jun;44:121-129 [FREE Full text] [doi: 10.1016/j.drugpo.2017.05.007] [Medline: 28578250]

22. Soussan C, Kjellgren A. The users of Novel Psychoactive Substances: online survey about their characteristics, attitudes and motivations. Int J Drug Policy 2016 Jun;32:77-84 [FREE Full text] [doi: 10.1016/j.drugpo.2016.03.007] [Medline: 27184218]
23. Soussan C, Andersson M, Kjellgren A. The diverse reasons for using Novel Psychoactive Substances - a qualitative study of the users' own perspectives. Int J Drug Policy 2018 Feb;52:71-78 [FREE Full text] [doi: 10.1016/j.drugpo.2017.11.003] [Medline: 29241144]
24. Research Standards. Bluelight. 2019 May 24. URL: https://www.bluelight.org/xf/pages/Research_Standards/ [accessed 2022-05-06]
25. Charmaz K. Teaching theory construction with initial grounded theory tools: a reflection on lessons and learning. Qual Health Res 2015 Dec;25(12):1610-1622. [doi: 10.1177/1049732315613982] [Medline: 26646825]
26. Strauss AL. Qualitative Analysis: For Social Scientists. New York, NY, USA: Cambridge University Press; 1987.
27. Strauss AL, Corbin JM. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. Thousand Oaks, CA, USA: Sage Publications; 1988.
28. Peacock A, Larance B, Farrell M, Cairns R, Buckley N, Degenhardt L. Opioid use and harms associated with a sustained-release Tapentadol formulation: a postmarketing study protocol. BMJ Open 2018 Mar 23;8(3):e020006 [FREE Full text] [doi: 10.1136/bmjopen-2017-020006] [Medline: 29574444]
29. Substance Abuse and Mental Health Services Administration. Key substance use and mental health indicators in the United States: results from the 2016 National Survey on Drug Use and Health. Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration. 2017. URL: https://www.samhsa.gov/data/sites/default/files/NSDUH-FFR1-2016/NSDUH-FFR1-2016.htm [accessed 2022-05-06]
30. Channell JS, Schug S. Toxicity of Tapentadol: a systematic review. Pain Manag 2018 Sep 01;8(5):327-339. [doi: 10.2217/pmt-2018-0027] [Medline: 30079795]
31. Gressler LE, Hammond DA, Painter JT. Serotonin syndrome in Tapentadol literature: systematic review of original research. J Pain Palliat Care Pharmacother 2017;31(3-4):228-236. [doi: 10.1080/15360288.2017.1416440] [Medline: 29336654]
32. Blanco C, Wall MM, Okuda M, Wang S, Iza M, Olfson M. Pain as a predictor of opioid use disorder in a nationally representative sample. Am J Psychiatry 2016 Dec 01;173(12):1189-1195. [doi: 10.1176/appi.ajp.2016.15091179] [Medline: 27444794]
33. Jann M, Kennedy WK, Lopez G. Benzodiazepines: a major component in unintentional prescription drug overdoses with opioid analgesics. J Pharm Pract 2014 Feb;27(1):5-16. [doi: 10.1177/0897190013515001] [Medline: 24436437]
34. Paulozzi LJ, Kilbourne EM, Shah NG, Nolte KB, Desai HA, Landen MG, et al. A history of being prescribed controlled substances and risk of drug overdose death. Pain Med 2012 Jan;13(1):87-95. [doi: 10.1111/j.1526-4637.2011.01260.x] [Medline: 22026451]
35. Stein MD, Kanabar M, Anderson BJ, Lembke A, Bailey GL. Reasons for Benzodiazepine use among persons seeking opioid detoxification. J Subst Abuse Treat 2016 Sep;68:57-61 [FREE Full text] [doi: 10.1016/j.jsat.2016.06.008] [Medline: 27431047]
36. Highlights of prescribing information - NUCYNTA®. U.S. Food & Drug Administration. 2008. URL: https://www.accessdata.fda.gov/drugsatfda_docs/label/2010/022304s003lbl.pdf [accessed 2022-05-06]
37. Cicero TJ, Ellis MS, Kasper ZA. Relative preferences in the abuse of immediate-release versus extended-release opioids in a sample of treatment-seeking opioid abusers. Pharmacoepidemiol Drug Saf 2017 Jan;26(1):56-62. [doi: 10.1002/pds.4078] [Medline: 27594167]

Abbreviations

ER: extended-release
IR: immediate-release
NMU: nonmedical use
TUIS: Tapentadol Use Internet Survey

Edited by A Mavragani; submitted 11.11.19; peer-reviewed by J Corkery, S Matsuda, H Demir; comments to author 27.01.22; revised version received 19.03.22; accepted 10.04.22; published 10.06.22.

Please cite as:
Vosburg SK, Dailey-Govoni T, Beaumont J, Butler SF, Green JL
Characterizing the Experience of Tapentadol Nonmedical Use: Mixed Methods Study
JMIR Form Res 2022;6(6):e16996
URL: https://formative.jmir.org/2022/6/e16996
doi:10.2196/16996
PMID:35687397

Original Paper

Feasibility and Acceptability of Internet-Based Interpersonal Psychotherapy for Stress, Anxiety, and Depression in Prenatal Women: Thematic Analysis

Katherine S Bright[1,2], BSc, BN, MN, PhD; Scott Stuart[1], MD; Deborah A Mcneil[3], PhD; Lindsay Murray[1], BMus; Dawn E Kingston[1], RN, PhD

[1]Faculty of Nursing, University of Calgary, Calgary, AB, Canada

[2]Women's Mental Health Clinic, Foothills Medical Centre, Alberta Health Services, Calgary, AB, Canada

[3]Maternal Newborn Child & Youth Strategic Clinical Network, Alberta Health Services, Calgary, AB, Canada

Corresponding Author:
Katherine S Bright, BSc, BN, MN, PhD
Faculty of Nursing
University of Calgary
PF2500B, 2500 University Drive NW,
Calgary, AB, T2N 1N4
Canada
Phone: 1 403 220 6262
Email: ksbright@ucalgary.ca

Abstract

Background: Prenatal mental health is a global health concern. Despite the far-reaching impact of prenatal mental health issues, many women do not receive the psychological care they require. Women in their childbearing years are frequent users of the internet and smartphone apps. Prenatal women are prime candidates for internet-based support for mental health care.

Objective: This study aimed to examine the feasibility and acceptability of internet-based interpersonal psychotherapy (IPT) for prenatal women.

Methods: Semistructured interviews were conducted with women who had received internet-based IPT modules with guided support as a component of a randomized controlled trial evaluating the scale-up implementation of a digital mental health platform (The Healthy Outcomes of Pregnancy and Postpartum Experiences digital platform) for pregnant women. Qualitative thematic analysis was used to explore and describe women's experiences. Data were analyzed for emerging themes, which were identified and coded.

Results: A total of 15 prenatal women were interviewed to examine their experiences and views on the feasibility and acceptability of internet-based IPT modules. Participants found the content informative and appreciated the ways in which the digital mental health platform made the IPT modules accessible to users. Participants voiced some differing requirements regarding the depth and the way information was presented and accessed on the digital mental health platform. The important areas for improvement that were identified were acknowledging greater depth and clarity of content, the need for sociability and relationships, and refinement of the digital mental health platform to a smartphone app.

Conclusions: This study provides useful evidence regarding treatment format and content preferences, which may inform future development. It also provides research data on the feasibility and acceptability of web-based applications for prenatal mental health care.

Trial Registration: ClinicalTrials.gov NCT01901796; https://clinicaltrials.gov/ct2/show/NCT01901796

(JMIR Form Res 2022;6(6):e23879) doi:10.2196/23879

KEYWORDS

internet-based; interpersonal psychotherapy; mental health; prenatal; anxiety; depression; stress; mobile phone

Introduction

Background

Prenatal mental health is a global health issue, with 15% to 25% of pregnant women experiencing clinical levels of depression, anxiety, and stress [1-3]. Left untreated, prenatal psychological distress is associated with a range of negative consequences on obstetrical outcomes, maternal functioning, infant and child development, interpersonal conflict, parenting strain, and postpartum mood disorders [4-7]. Despite the far-reaching impact of prenatal mental health issues and well-established recommendations for routine mental health screening during the prenatal period, <20% of providers routinely screen [8]. Equally concerning is that only 1 in 7 prenatal women obtain the mental health intervention they require [8,9].

Interpersonal psychotherapy (IPT) is a highly effective treatment for depression and anxiety [10-12]. IPT improves symptoms of depression and anxiety in prenatal women [13-16]. Stuart and O'Hara [17] suggested that IPT is a mainline treatment for prenatal mental health as it focuses on addressing 4 interpersonal problems: role transitions, interpersonal disputes, grief, and interpersonal deficits. These 4 areas address the significant factors involved in the prediction and maintenance of depression and anxiety in pregnant women. The US Preventive Task Force has identified convincing evidence that IPT is effective in treating prenatal depression [18]. Although IPT is considered an effective treatment for depressive and anxiety disorders in the prenatal period, many women are uncertain about what is considered anticipated or expected mental health experiences during pregnancy, and as such, many women are reluctant to access mental health care [10,13,19-23]. Additional barriers to access to face-to-face IPT remain a challenge because of limited IPT-trained therapists, long wait times to access care, and the high cost of therapy sessions [24,25]. As a result, there is a need to make effective IPT interventions more readily available and accessible to pregnant women.

Internet-Based Interventions for Prenatal Mental Health

Internet-based interventions are ideal treatment options for prenatal women to overcome major obstacles to accessing therapy, such as long wait times, busy schedules, stigmas to accessing care, and the financial burden associated with treatment [24]. In addition, internet-based interventions demonstrate preliminary effectiveness in the prevention [26,27] and treatment [28,29] of prenatal depression. Guided internet-based treatments have high levels of patient adherence and convincing reductions in mental health symptoms. They are readily available and cost-effective alternatives to face-to-face treatment [25,30]. Guided support, as a component of an internet-based treatment, has been reported in several systematic reviews to increase adherence and effectiveness for participants [31-33]. Guided support permits personal connections when needed and is a beneficial feature of internet-based interventions [34,35].

The Healthy Outcomes of Pregnancy and Postpartum Experiences With the IPT Program

We developed an internet-based prenatal mental health intervention for a low-intensity, guided support IPT program for stress, anxiety, and depression. The IPT program comprises 6 modules adapted from the IPT clinician guide developed by Stuart and Robertson [36] and tailored to the specific needs of prenatal women. These IPT modules were delivered through a digital mental health platform (The Healthy Outcomes of Pregnancy and Postpartum Experiences [HOPE] digital platform). Given the model of IPT developed by Stuart and Robertson [36] as benefiting problems involving role transitions, loss, and interpersonal conflict, 6 modules were built around these 3 areas [20]. Role transitions during the prenatal and postpartum periods are primarily related to developing new skills and accommodating changing responsibilities while maintaining relationships. This problem area explores the multiple roles women juggle and the increased relationship demands because of these roles [17,36]. Women are encouraged to combine new roles rather than give up old roles, express emotions attached to each of the roles and their impact on self-esteem, and explore ambivalent feelings for each role. Ultimately, the intention of exploring role transitions is to assist women in developing a more balanced understanding of each role, modifying expectations, and assisting them in restructuring priorities [17,36].

The focus area of loss explores grief reactions that coincide with pregnancy and the arrival of a newborn [36]. The therapeutic intent of exploring loss is to assist with the mourning process and help women cultivate new and current relationships that can be substituted for relationships that were lost [17,36]. The focus area of interpersonal conflict is one of the most significant possible stressors during pregnancy and into the postpartum period. This often occurs in individuals and their spouses [17,36].

This internet-based prenatal mental health intervention comprises six 30-minute, web-based IPT modules delivered over 6 to 8 weeks. The topics of the modules are (1) identifying the important relationships in women's lives, (2) understanding and improving communication patterns, (3) navigating interpersonal disputes, (4) adapting to role transitions, (5) working through grief and loss, and (6) maintaining IPT strategies and carrying these skills forward in their lives after the study ends. These topics were tailored to pregnant women. Participants were asked to complete specific assignments on the web for each module, such as self-awareness homework and exploring relationships with those close to them. The goals of the intervention included symptom relief, improving interpersonal functioning and relationships, changing expectations about interpersonal relationships, and improving social support networks. Participants were guided to recognize and disengage from unhelpful communication patterns and foster strategies for developing and nurturing social support in navigating challenging times, such as role transitions, conflict with their partner or extended family members, and grief or loss.

Within this program, women were encouraged to assess their expectations for roles that they, their partners, parents, in-laws, and other children have during the prenatal period. Within these expectations, women explore the changes and consistency of roles before, during, and after pregnancy into the postpartum period. A significant aspect of the focus area of interpersonal conflict includes the identification of disputes and the development of problem-solving approaches that women can put into action. In addition, this IPT module contained the following components: development of a support system, effective communication strategies, and skills for managing conflict in relationships [36]. Homework exercises involved assessing attachment style; communication style; breaking down a distressing conversation or interaction; understanding one's relationships through relationship circles; visualizing, describing, and resolving conflict through a disagreement or dispute graph; and understanding role transitions through a life events timeline. Women were assessed for the areas of life challenges that caused emotional distress. This provided a targeted direction regarding modules that best suited their needs (eg, loss, transition, and interpersonal).

The IPT internet-based prenatal mental health intervention, comprising 6 IPT modules, was a component of a randomized controlled trial that evaluated the scale-up of a digital mental health platform for pregnant women—the HOPE digital mental health platform.

This study aimed to investigate women's views of the feasibility and acceptability of IPT internet-based prenatal mental health modules delivered through a digital mental health platform with low-intensity, guided support.

Methods

Overview

Semistructured interviews were conducted in a flexible and responsive manner as a method of data collection. Qualitative content analysis and thematic analysis were used as methods of data analysis. These qualitative interviews sought to understand women's individual experiences of interacting with the digital mental health platform and address questions regarding the feasibility and acceptability of the internet-based IPT intervention.

Design

A qualitative design was used to assess women's individual experiences regarding the feasibility and acceptability of internet-based IPT interventions. This study adopted a pragmatic approach.

Recruitment

A total of 20 participants in the intervention group were contacted by email to assess their interest in participating in an interview and then followed up by email to set up a suitable time for the telephone interview. Of the 20 participants, 15 (75%) women in the intervention group agreed to be interviewed by a member of the research team (KSB). Before the interviews, information on computer-generated program use was collected. Participants who completed more than half of the modules were considered *high users*, and those who completed less than half of the modules were considered *low users*. We also identified participants' levels of engagement and explored the barriers to and facilitators of engagement.

Data Collection

Semistructured interviews, lasting 45 to 75 minutes, were conducted in a flexible and responsive manner as a method of data collection. These interview questions were piloted with 3 prenatal women who were not in the study to ensure that the questions were clear and concise. The interviews addressed key areas based on guidelines for assessing internet intervention research [37]: assessment and navigation of the intervention, acceptability and perceived usefulness of various components of the intervention, and recommendations for improvement [38,39]. The interview questions were used as a guide; however, the interviews were conducted with the flexibility to allow participants to freely discuss topics. Recruitment continued during data analysis to achieve saturation of the themes. All interviews were conducted over the phone, audio recorded, and conducted by the first author (KSB; she or her), an experienced qualitative research interviewer, as a part of her doctoral dissertation and transcribed verbatim. KSB is a clinician in an outpatient reproductive psychiatric clinic for the past 10 years. Field notes were made following each interview. Participants received a CAD $5 (US $4) coffee card as compensation for their time.

Analysis

Transcripts were coded and analyzed using thematic analysis techniques to identify categories and themes [40,41]. The approach used for this thematic analysis was guided by the method described by Braun and Clarke [40]. The semistructured interview guide provided a focused direction for the interviews. The themes that concurrently emerged from these interviews were relevant to the aim of our study and our research questions. A coding framework was developed in response to the themes that emerged during each interview. This coding framework was reviewed and refined as it was applied to data. Patterns within and across themes were explored using an analysis process. The main coding categories paralleled the questions asked during the interviews. Categories were also reflective of emerging trends in the data, which became apparent from the frequency of certain categories and repetition of points. Agreement on the categories and concepts was sought between members of the research team to ensure reliability. The interviews and coding framework were examined until no new information emerged from the data. KB, who has a background in qualitative interviewing, conducted the interviews and was the lead in the data analysis. LM and DEK, who have backgrounds in psychology, mental health, and the prenatal population, each read 5 interview transcripts. The coding framework was discussed throughout its development, with regular meetings between the 3 researchers to ensure that the concepts were appropriately identified and described.

Ethics Approval

This study was registered with ClinicalTrials.gov (NCT01901796) and was approved by the University of Calgary Research Ethics Board (REB16-0061 U of Calgary).

Results

Overview

Of the 20 approached women, 15 (75%) agreed to participate in the interviews. Of the 15 participants, 10 (67%) were considered high users of the IPT modules, and the remaining 5 (37%) were classified as low users. The women ranged in age from 28 to 38 years. All women were partnered, and 67% (10/15) of women had ≥1 child.

The themes identified were feasibility and acceptability. Women reported the feasibility of the digital mental health platform (the HOPE platform) and IPT modules in 3 subthemes: treatment feasibility, flexible access, and impact of mental health status on platform engagement. Treatment feasibility refers to the ease with which the platform was integrated into women's lives. Acceptability also had 3 subthemes: clarity and depth of content, lack of relationships with the internet-based IPT, and suggestions for improvement.

Feasibility

Overall, women indicated that the delivery of internet-based IPT modules through the HOPE digital mental health platform was a way for them to self-manage their mental health during their pregnancy. Participants reported that they were pleased to find a mental health resource specifically designed for pregnant women. IPT modules and the platform were easily accessible at times and places that were convenient to the participants.

Treatment Feasibility

Overall, the women reported that the features of the internet-based IPT modules and platform appeared at the right time in their lives to be easily integrated into their lives and practices. Women reported that they wanted to participate in this internet-based IPT study as it was an opportunity to access support and information. They also reported an interest in accessing resources to assist with the transition through pregnancy and into the postpartum period:

> *I was looking for an independent way to learn more about what I was going through in pregnancy.* [P1]
>
> *I have mental health issues so I knew that it would be a good support for me and I just want to find out more information and more resources to just help me with this transition time, with pregnancy. It was, just kind of encouraging to see, the support there.* [P2]
>
> *I do know myself pretty well and know that I have a tendency to be more of an anxious person so I thought this might be a good opportunity to be proactive with my mental health.* [P6]
>
> *I think the biggest thing was that it was a resource specifically targeted to exactly where I was in my life, just in the middle of my pregnancy, kind of near the end, things that you are thinking about at that time mental health wise. It was just clearly targeted for where I was at.* [P8]
>
> *I joined this study because I have a history of postpartum depression with my first daughter so I figured I could use all the support I could find with my next pregnancy...I thought that it is good, we need more research into that.* [P11]
>
> *I had experienced depression one time before and I was looking for a resource that would help prevent postpartum depression. This program seemed hopeful. It targeted the time in pregnancy and that was exactly where I was at.* [P14]

Flexible Access

Participants enjoyed that the internet-based IPT modules could be accessed from home or at a location and time that suited them. This was reported as important as most women (14/15, 93%) were either working outside the home and inside the home with one or more children:

> *I'm actually able to do more exercises and modules after work.* [P5]
>
> *It was the perfect amount of time to complete them. Like you could logon on your lunchbreak and go through a module. That is what I did.* [P6]
>
> *I completed the modules and surveys on my phone. It was easy to do anywhere when I had the time...If you want to make the time for it you will do it.* [P11]
>
> *Web-app made it easy to use and access from anywhere.* [P12]
>
> *I quite like being in the comfort of my own home, logging on when I can if I can...I do like that they (the modules) were short and sweet...I found that shortness of the modules was really important for me to start and finish and get something out of it rather than just get part way through it and have to stop.* [P13]

Impact of Mental Health Status on Platform Engagement

A few participants reported that their mental health status at the moment predicted their engagement on the platform and their motivation to work through the modules. When they were doing well mentally, they reported that they were able to work through the internet-based IPT modules. When they felt they were not doing well, they did not access the platform modules:

> *When I am good, I'm good, right? When I'm bad, it's the complete opposite. That I totally shut down and that's when you need something the most.* [P4]
>
> *I have so much stuff going on right now, I am 6 months pregnant, working, and feeling really anxious. To tell you the truth I kind of forgot about the Web-App and modules. It has been a couple of months and I think it would have been helpful to have a reminder every day to cue me back to it.* [P5]
>
> *It was hard to find the time to really pay attention and I think I was also a little bit too deep into my mental health issues at that time to really be able to*

appreciate the app and modules for what they were. [P5]

Acceptability

Overview

Participants reported that they found the digital mental health platform (the HOPE platform) and IPT modules user-friendly and helpful for increasing their skills to solve real-life concerns. The participants greatly appreciated the guided support or coaching aspect of the digital platform. Most of the mental health platform users reported that they would have preferred the digital platform to be a smartphone app—an app that can be accessed on their phone rather than having to access the IPT modules through a browser over the internet. Many participants reported that the modules contained information that was relevant to them, and they wished that the modules had been developed in greater depth. Although participants enjoyed the HOPE digital mental health platform and IPT modules, they missed the person-to-person connections that occur with individual and group therapy.

Usefulness

Overall, participants reported that the internet-based IPT modules were acceptable, displayed well on their smartphones or computers, and were user-friendly and modern:

> *I actually appreciated that they were very easy to interact with. I didn't feel like I was lost looking for buttons or cues or having to read things again.* [P1]
>
> *I thought the interface was pretty clear and pretty easy to interact with. I could record my answers quite easily.* [P2]
>
> *The interface of the app was easy to use.* [P9]

Women reported that the content in the modules was useful, relevant for their current needs, increased their self-awareness, and helped them solve real-life problems:

> *It was very useful actually. I've never sat down and really understood how my tendencies would impact the way that I was interacting with my husband specifically.* [P1]
>
> *I think one of the most import parts of the modules was communication skills and being assertive...I liked how it gave you an example of how it could help in real life...I am keeping in mind some of the communication pieces as we are navigating some of the postpartum issues.* [P6]
>
> *I liked how the modules really made you think about what is going on for you. They are definitely teaching you things that make you take a different look at what is going on in your life.* [P7]
>
> *I think that the biggest benefit of the modules was just a different way of thinking of situations differently or providing coping tools to address the here and now...I definitely found them helpful.* [P9]
>
> *I am currently having some issues in a friendship, and I think that the relationship module is going to be really helpful in resolving this.* [P11]
>
> *I really liked that I could have a project to work on, something a little more concrete to work though because having something concrete is helpful for me to learn.* [P12]

Many participants reported the benefits of the guided support or coach component of the digital mental health platform:

> *I liked the coach. That was probably my favorite part. It's just having that someone that just checked in...it's just that extra check and like I said, you know, if you're not comfortable going to a friend.* [P4]
>
> *I think that the biggest benefit to me was when I put down the need for a call, somebody actually did call me...It was nice to just have somebody on the other end of the line who was like we saw that you wanted to be called by a coach...It was really nice to have someone reach out and see how I was doing and to see what resources I needed.* [P9]
>
> *I talked to the coach for about 45 minutes, and it was really helpful. She did a fantastic job of talking with me about what was going on for me.* [P12]
>
> *I was contacted by the coach and that was really good. I didn't need any resources or anything, but it was nice to have her check in on me. I found that really helpful.* [P14]
>
> *I had a call from the coach, and I found that more helpful...Just having someone to talk who already knows I am pregnant and already knows that I am in a very vulnerable state during pregnancy, and they knew the app better than I did and they could tell me where to go or what local resources are available.* [P15]

All participants stated that they would prefer the IPT modules to be delivered through a smartphone app rather than the web application:

> *The one thing would be more accessible would be if it was just an app that you can download. To use on the go.* [P2]
>
> *I think a mobile app would really help. Like I don't really even check my email anymore. I think a true app would really help with that.* [P6]
>
> *I would have preferred to see this in an app form instead of a Web-App.* [P8]
>
> *It was a little disappointing that is wasn't just an app that I could find on my phone and that I had to log on through a web browser.* [P10]
>
> *I thought it was an App, but it was a Web-App. That was a bit misleading. Either change the name to Web-App or make it an App.* [P11]

Clarity and Depth of Content

Of the 15 participants, 12 (80%) reported that the content of the IPT modules required further refinement. These participants reported that all topics were relevant; however, they wanted to go further, recommending a greater depth of material:

> *It was nothing new, I kind of knew, like, information about attachment style like its nothing I haven't heard before. So, it was good there were a couple of examples of case studies, Yeah, like I said but it's very, kind of, basic.* [P2]
>
> *It wasn't necessarily new information.* [P4]
>
> *I think for me sometimes it was a bit generic or really surface level and I understand that in these modules you can't really get into the nitty gritty but maybe it is because of my background that I was like I kind of knew that.* [P6]
>
> *I felt like the modules were really short, so you kind of got into a topic, and some of them were really useful and helpful, but they seemed really brief. Like they could have gone into more detail.* [P8]
>
> *I think if you had a list of additional resources in the modules for more information if you need it or if you really wanted to go further.* [P10]
>
> *The introductory ones seemed ok but pretty basic. But the one that I have stopped at, the one that has an exercise, and I am hoping that the future ones are more like that because it seems like they will actually be more useful and have new tools for me.* [P11]
>
> *I felt there the information was good but there was nothing like to go from here. Like say with the attachment style one, I really liked learning about attachment styles, but it didn't really give me, and I wanted more information like this is how your approach your partner and this is how your partner approaches you. Like I wanted to know what I was and then I wanted to have tools and I felt like I didn't that I was given the toolbox but not the tools.* [P12]
>
> *A lot of the information was not new to me but the reminder of it is still really important.* [P13]

Many participants reported that the modules required further refinement of the language used throughout and clarity of instructions for exercises or homework:

> *The verbiage in a couple of the modules do have the inclination, to be that way is bad and to be this way is the better option...to do an honest assessment of where they are at, its understanding that your communication style and your attachment style is never bad it's just where you are right now.* [P3]
>
> *I think that is the area where I struggled most. Sometimes I wasn't even clear what I was supposed to do or what the homework was.* [P6]
>
> *That's the part that I am stuck on now. I remember not being able to fully understand it. And I think that is a big reason why I wasn't able understand what I could take from what is going on in my life right now and to be able to implement it to complete the exercise. I needed a better understanding of what the homework was asking because I wasn't understanding what it was asking.* [P7]
>
> *I think that if the app could have it so you could do a lot of the exercises in real time so actually inputting your thoughts and your data into app and it could organize it for you. I think that would make it a lot more attractive to the user.* [P9]
>
> *I think clearly outlining what is available on this Web-App from the beginning would help like the surveys, the modules, the mood tracker, and things like that. And giving people an idea of what is going to happen when so that it is not forgotten or confusing when things come up.* [P10]
>
> *I have competed over half of the modules, but I am stuck on one because it asks you to do a homework assignment outside of the app and I am stuck. I don't want to blow past in and not do it. I need to sit down and make the time to do it. It just hasn't happened yet.* [P11]
>
> *I think that the modules could focus on how exactly to ask for help from others.* [P13]

Personal Connection

Participants reported that although they enjoyed the ease of the internet-based IPT modules and screening, they missed the opportunity to connect with other individuals:

> *I remember just thinking I needed more and that is where the personal sessions would help a lot more.* [P11]
>
> *I am a big advocate for people helping people. I appreciate the ease of online work but also miss the face-to-face interaction you get with a therapist...Maybe online you need to offer both for people to have the option to do it on their own or to login into a closed group if they want, appealing to different personalities.* [P13]
>
> *It is hard for me to properly put down a number indicating how I feel because it is based on the day, so I just like it was very generic and I am not really sure if it got, understood where I was coming from. It was more helpful for me once I got a hold of someone to talk to. It was the personal connection that was important to me...Technology is great, but it can only do so much.* [P15]

Suggestions for Improvement

Suggestions for improvement for many women included incorporating different learning styles:

> *I find that I learn a lot by watching things too. Like there is so much on YouTube and TedTalks to find helpful, like Brené Brown or something cool like that would be something for interpersonal relationships and vulnerability...It is about catering to different learning styles.* [P2]
>
> *I am a very visual learner so having videos instead of a lot of reading would have been more helpful for me.* [P7]
>
> *I used multiple apps during my pregnancy it would have been nice to have one that tracks where I am in my pregnancy and mental health to see if they could*

> *be linked or incorporated into one app would be something that I would be interested in.* [P8]
>
> *Having the modules as an audio, like something I could listen to while driving, might be helpful.* [P15]

In addition, the participants suggested embedding mental health into a pregnancy app to include mental health as a part of overall health:

> *I think it would be helpful to integrate this app into a pregnancy app, like all of us had the bump app, and it kind of goes through some of the physical changes in pregnancy and I think people are kind of curious about that the baby is doing and what their bodies are doing, so that might be nice to tie together. That is more holistic too. So, showing people that mental health is just health.* [P6]

In addition, there was a suggestion to extend the IPT work into the postpartum period with modules aimed at the challenges of being a new parent:

> *I think it would be helpful to have additional modules to work through in the postpartum period. Like navigating yourself, your new baby, and your relationship with your partner in the postpartum period.* [P12]

Overall, the participants found the HOPE digital mental health platform and IPT modules helpful in managing their mental health concerns during pregnancy. They also had many suggestions to further improve the HOPE digital mental health platform and IPT modules. These women suggested that the platform and modules could attend to a variety of learning styles, including the use of videos and audio components. In addition, the participants wanted this mental health component to be integrated into existing pregnancy apps as part of comprehensive prenatal and postpartum care.

Discussion

Principal Findings

This study aimed to assess the feasibility and acceptability of a pilot study of internet-based IPT modules. Regarding the feasibility of the study, participants reported that the features of the internet-based IPT modules delivered via the digital mental health platform (the HOPE digital platform) were easily integrated into their lives. Of the 15 participants, 10 (66%) were described as high users of the modules, completing >4 of the 8 modules. The internet-based delivery of the IPT modules permitted flexible and independent access. Participants appreciated being able to seek support and resources to self-manage the emotional or mental health challenges arising during their pregnancies.

Participants reported that the format of the digital mental health platform was acceptable and user-friendly. They also reported that the digital mental health platform could be improved by delivering the format through a smartphone app. Among the women who accessed the internet-based IPT, the treatment was considered generally useful and helpful in examining communication and interactions with others. Many women reported that the modules were helpful in solving concrete problems. Although described as useful, there were elements of the modules that the participants felt could be improved. Some considered the information to be too basic and required further refinement. They also sought improvement in the depth and breadth of content and clarity of instructions for exercises and homework. In addition, the participants suggested that they would benefit from additional directions on what they needed to do when not managing well.

Participants reported that they appreciated the guided support or coach aspect of the internet-based IPT modules and digital mental health platform. Whether the coach was triggered in the digital mental health platform to contact women (based on responses to survey questionnaires) or the women requested a coach callback, the guided support or coach gave women the feeling that they were *being checked in on*. They valued having the coach see how they were doing. Participants reported that this was one of the most important aspects of the IPT modules and digital mental health platform.

Comparison With Prior Work

This study found that an internet-based mental health intervention was feasible, with participants citing the ease and flexibility of access as beneficial. These results are consistent with previous studies [42-45]. As mentioned by the participants in this study, a smartphone app is preferred over the delivery of interventions with a web application. Previous studies have also reported an increase in treatment accessibility and participation in mental health interventions when delivered via smartphone apps [46]. Furthermore, internet-based and smartphone apps provide convenient and potentially anonymous access to treatment. Smartphone apps can also overcome conventional barriers to seeking help, including lack of time, stigma, childcare issues, and embarrassment [47-49].

The efficacy of internet-based interventions is well-established for various mental health disorders, including depression, anxiety, agoraphobia, panic disorder, and stress [25,50,51]. Acceptability of internet-based interventions has been reported across different populations, including students [52], perinatal women [53], and older adults [54].

This study adds to the evidence that pregnant women seek resources and support to self-manage mild to moderate mental health concerns [55]. Consistent with previous studies, the preference of women to address mental health problems on their own highlights the need and opportunity to offer alternatives to traditional face-to-face, therapist-led interventions [56-58]. Providing prenatal women with effective resources and support to self-manage their mild to moderate mental health concerns, including internet-based therapy, may be an effective approach for pregnant women with mild to moderate mental health concerns.

This study was unique in that mental health symptoms may impact engagement in the digital mental health platform. Participants in this study reported that they were less inclined to access the web application platform when they were doing very well and when they were struggling with significant psychological distress. This finding seems to contradict the

findings that mental health symptoms had no relationship with a preference for and engagement in web-based mental health programs [59,60]. These studies indicate that neither high nor low symptoms of depression and anxiety caused individuals to be more or less likely to engage in web-based mental health interventions.

Guided support is a beneficial feature of internet-based interventions; this preference for person-to-person support is consistent with previous work [34,35]. However, evidence regarding the nature of guided support is lacking. The use of face-to-face or person-to-person mental health programs is slightly preferred for those experiencing emotional and personal problems [61-64]. Predictors for the preference of person-to-person mental health interventions included currently experiencing emotional or personal problems. The use of coach-based guided support along with mental health apps for smartphones has been found to improve mental and emotional health by decreasing participants' depression, anxiety, and general distress [46,65-67]. Guided support or coaching with internet-based mental health interventions has been reported to be an effective, cost-efficient, and acceptable alternative to face-to-face therapy [33]. Systematic reviews exploring the internet-based mental health interventions found that offering some form of support or guidance during the web-based treatment increased the effectiveness and was associated with higher levels of program completion [34,68]. Findings from this study will help move this area of study forward by identifying the importance of coach-based guided support within a pregnant population. This study may lead to an increase in the refinement of smartphone delivery of IPT in pregnant women.

Limitations

The findings of the usability, feasibility, and acceptability of internet-based IPTs are limited to the experiences and perspectives of a small sample. None of the women who withdrew from the study before completing the first module reported reasons for dropping out of the program.

The research team took several actions to minimize the likelihood of bias during the analysis phase. The research team had regular discussions about the coding framework with fellow team members who each reviewed 5 interviews to ensure that the concepts were being appropriately identified and described.

Participants who agreed to be interviewed were offered a small monetary compensation for their time, which may have affected their responses.

Another limitation may be the generalizability of these findings because of the smaller prenatal sample size.

Clinical Implications

This study provides insights into the acceptability of this internet-based IPT program for prenatal women. This information could be used to further refine the development of targeted IPT mental health support for women during their pregnancy and into the postpartum period. Smartphone app development is required for the delivery of IPT programs. This will support women during their pregnancy and into the postpartum period. Future smartphone IPT programs aimed at prenatal women would also benefit from guided or coach support components.

Conclusions

This study demonstrated that there is a significant interest in prenatal mental health internet-based IPT treatment. The current format of this internet-based IPT needs refinement. This study revealed the strengths and weaknesses of the content and format of this internet-based IPT. In addition, the study participants highlighted potential areas of improvement. The IPT modules for women during the prenatal period would benefit from the further refinement of the content and integration of the revised modules into a randomized controlled trial to explore the efficacy of the internet-based IPT modules on perinatal mental health concerns.

Authors' Contributions

KSB conducted qualitative interviews. KSB, LM, and DEK analyzed qualitative data. All the authors (KSB, DAM, LM, SS, and DEK) participated in the refinement of the study methods, critically reviewed, and provided feedback on the final version submitted for publication, in accordance with the International Committee of Medical Journal Editors criteria.

Conflicts of Interest

None declared.

References

1. Goodman JH, Tyer-Viola L. Detection, treatment, and referral of perinatal depression and anxiety by obstetrical providers. J Womens Health (Larchmt) 2010 Mar 19;19(3):477-490. [doi: 10.1089/jwh.2008.1352] [Medline: 20156110]
2. Khan L. Falling through the gaps: perinatal mental health and general practice. Centre for Mental Health. URL: https://www.centreformentalhealth.org.uk/sites/default/files/2018-09/falling.pdf [accessed 2022-05-14]
3. O'Hara MW, Wisner KL. Perinatal mental illness: definition, description and aetiology. Best Pract Res Clin Obstet Gynaecol 2014 Jan 28;28(1):3-12 [FREE Full text] [doi: 10.1016/j.bpobgyn.2013.09.002] [Medline: 24140480]
4. Tomfohr L, Buliga E, Letourneau N, Campbell T, Giesbrecht GF. Trajectories of sleep quality and associations with mood during the perinatal period. Sleep 2015 Aug 01;38(8):1237-1245 [FREE Full text] [doi: 10.5665/sleep.4900] [Medline: 25845691]

5. Kozyrskyj AL, Letourneau NJ, Kang LJ, Salmani M. Associations between postpartum depressive symptoms and childhood asthma diminish with child age. Clin Exp Allergy 2017 Mar;47(3):324-330. [doi: 10.1111/cea.12837] [Medline: 27770463]
6. Lebel C, Walton M, Letourneau N, Giesbrecht GF, Kaplan BJ, Dewey D. Prepartum and postpartum maternal depressive symptoms are related to children's brain structure in preschool. Biol Psychiatry 2016 Dec 01;80(11):859-868. [doi: 10.1016/j.biopsych.2015.12.004] [Medline: 26822800]
7. Letourneau NL, Dennis CL, Benzies K, Duffett-Leger L, Stewart M, Tryphonopoulos PD, et al. Postpartum depression is a family affair: addressing the impact on mothers, fathers, and children. Issues Ment Health Nurs 2012 Jul;33(7):445-457. [doi: 10.3109/01612840.2012.673054] [Medline: 22757597]
8. Kim JJ, La Porte LM, Corcoran M, Magasi S, Batza J, Silver RK. Barriers to mental health treatment among obstetric patients at risk for depression. Am J Obstet Gynecol 2010 Mar;202(3):312.e1-312.e5. [doi: 10.1016/j.ajog.2010.01.004] [Medline: 20207252]
9. Bowen A, Bowen R, Butt P, Rahman K, Muhajarine N. Patterns of depression and treatment in pregnant and postpartum women. Can J Psychiatry 2012 Mar;57(3):161-167. [doi: 10.1177/070674371205700305] [Medline: 22398002]
10. Cuijpers P, Geraedts AS, van Oppen P, Andersson G, Markowitz JC, van Straten A. Interpersonal psychotherapy for depression: a meta-analysis. Am J Psychiatry 2011 Jun;168(6):581-592 [FREE Full text] [doi: 10.1176/appi.ajp.2010.10101411] [Medline: 21362740]
11. Donker T, Bennett K, Bennett A, Mackinnon A, van Straten A, Cuijpers P, et al. Internet-delivered interpersonal psychotherapy versus internet-delivered cognitive behavioral therapy for adults with depressive symptoms: randomized controlled noninferiority trial. J Med Internet Res 2013 May 13;15(5):e82 [FREE Full text] [doi: 10.2196/jmir.2307] [Medline: 23669884]
12. Stangier U, Schramm E, Heidenreich T, Berger M, Clark DM. Cognitive therapy vs interpersonal psychotherapy in social anxiety disorder: a randomized controlled trial. Arch Gen Psychiatry 2011 Jul;68(7):692-700. [doi: 10.1001/archgenpsychiatry.2011.67] [Medline: 21727253]
13. Spinelli MG. Interpersonal psychotherapy for depressed antepartum women: a pilot study. Am J Psychiatry 1997 Jul;154(7):1028-1030. [doi: 10.1176/ajp.154.7.1028] [Medline: 9210760]
14. Spinelli MG, Endicott J. Controlled clinical trial of interpersonal psychotherapy versus parenting education program for depressed pregnant women. Am J Psychiatry 2003 Mar;160(3):555-562. [doi: 10.1176/appi.ajp.160.3.555] [Medline: 12611838]
15. Brandon AR, Ceccotti N, Hynan LS, Shivakumar G, Johnson N, Jarrett RB. Proof of concept: partner-assisted interpersonal psychotherapy for perinatal depression. Arch Womens Ment Health 2012 Dec;15(6):469-480 [FREE Full text] [doi: 10.1007/s00737-012-0311-1] [Medline: 23053218]
16. Sockol LE, Epperson CN, Barber JP. A meta-analysis of treatments for perinatal depression. Clin Psychol Rev 2011 Jul;31(5):839-849 [FREE Full text] [doi: 10.1016/j.cpr.2011.03.009] [Medline: 21545782]
17. Stuart S. Interpersonal psychotherapy for postpartum depression. Clin Psychol Psychother 2012 Apr 02;19(2):134-140 [FREE Full text] [doi: 10.1002/cpp.1778] [Medline: 22473762]
18. US Preventive Services Task Force, Curry SJ, Krist AH, Owens DK, Barry MJ, Caughey AB, et al. Interventions to prevent perinatal depression: US Preventive Services Task Force Recommendation statement. JAMA 2019 Feb 12;321(6):580-587. [doi: 10.1001/jama.2019.0007] [Medline: 30747971]
19. Cuijpers P, van Straten A, Andersson G, van Oppen P. Psychotherapy for depression in adults: a meta-analysis of comparative outcome studies. Focus 2010 Jan;8(1):75. [doi: 10.1176/foc.8.1.foc75]
20. Donker T, Batterham P, Warmerdam L, Bennett K, Bennett A, Cuijpers P, et al. Predictors and moderators of response to internet-delivered Interpersonal Psychotherapy and Cognitive Behavior Therapy for depression. J Affect Disord 2013 Oct;151(1):343-351. [doi: 10.1016/j.jad.2013.06.020] [Medline: 23953024]
21. Markowitz JC, Lipsitz J, Milrod BL. Critical review of outcome research on interpersonal psychotherapy for anxiety disorders. Depress Anxiety 2014 Apr 03;31(4):316-325 [FREE Full text] [doi: 10.1002/da.22238] [Medline: 24493661]
22. Stuart S, Clark E. [The treatment of postpartum depression with interpersonal psychotherapy and interpersonal counseling]. Sante Ment Que 2008;33(2):87-104 [FREE Full text] [doi: 10.7202/019670ar] [Medline: 19370259]
23. Gennaro S, O Connor C, McKay E, Gibeau A, Aviles M, Hoying J, et al. Perinatal anxiety and depression in minority women. MCN Am J Matern Child Nurs 2020;45(3):138-144 [FREE Full text] [doi: 10.1097/NMC.0000000000000611] [Medline: 31977497]
24. Kingston D, Tough S. Prenatal and postnatal maternal mental health and school-age child development: a systematic review. Matern Child Health J 2014 Sep;18(7):1728-1741. [doi: 10.1007/s10995-013-1418-3] [Medline: 24352625]
25. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 2010 Oct 13;5(10):e13196 [FREE Full text] [doi: 10.1371/journal.pone.0013196] [Medline: 20967242]
26. Barrera AZ, Wickham RE, Muñoz RF. Online prevention of postpartum depression for Spanish- and English-speaking pregnant women: a pilot randomized controlled trial. Internet Interv 2015 Sep 01;2(3):257-265 [FREE Full text] [doi: 10.1016/j.invent.2015.06.002] [Medline: 26273567]

27. Haga SM, Drozd F, Brendryen H, Slinning K. Mamma mia: a feasibility study of a web-based intervention to reduce the risk of postpartum depression and enhance subjective well-being. JMIR Res Protoc 2013 Aug 12;2(2):e29 [FREE Full text] [doi: 10.2196/resprot.2659] [Medline: 23939459]
28. Danaher BG, Milgrom J, Seeley JR, Stuart S, Schembri C, Tyler MS, et al. Web-based intervention for postpartum depression: formative research and design of the MomMoodBooster program. JMIR Res Protoc 2012 Nov 22;1(2):e18 [FREE Full text] [doi: 10.2196/resprot.2329] [Medline: 23612274]
29. O'Mahen HA, Richards DA, Woodford J, Wilkinson E, McGinley J, Taylor RS, et al. Netmums: a phase II randomized controlled trial of a guided Internet behavioural activation treatment for postpartum depression. Psychol Med 2013 Oct 09;44(8):1675-1689. [doi: 10.1017/s0033291713002092]
30. Andrews G, Titov N. Is internet treatment for depressive and anxiety disorders ready for prime time? Med J Aust 2010 Jun 07;192(S11):S45-S47. [doi: 10.5694/j.1326-5377.2010.tb03693.x] [Medline: 20528709]
31. Spek V, Cuijpers P, Nyklícek I, Riper H, Keyzer J, Pop V. Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. Psychol Med 2007 Mar;37(3):319-328. [doi: 10.1017/S0033291706008944] [Medline: 17112400]
32. Titov N. Internet-delivered psychotherapy for depression in adults. Curr Opin Psychiatry 2011 Jan;24(1):18-23. [doi: 10.1097/yco.0b013e32833ed18f]
33. Andersson G, Cuijpers P, Carlbring P, Riper H, Hedman E. Guided internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: a systematic review and meta-analysis. World Psychiatry 2014 Oct;13(3):288-295 [FREE Full text] [doi: 10.1002/wps.20151] [Medline: 25273302]
34. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
35. Johansson R, Andersson G. Internet-based psychological treatments for depression. Expert Rev Neurother 2012 Jul;12(7):861-9; quiz 870. [doi: 10.1586/ern.12.63] [Medline: 22853793]
36. Stuart S, Robertson M, Robertson M. Interpersonal Psychotherapy 2E A Clinician's Guide. Boca Raton, FL, USA: CRC Press; 2012.
37. The Hope Project. Dr. Dawn Kingston. URL: https://www.drdawnkingston.com/the-hope-project/ [accessed 2022-05-30]
38. Proudfoot J, Klein B, Barak A, Carlbring P, Cuijpers P, Lange A, et al. Establishing guidelines for executing and reporting internet intervention research. Cogn Behav Ther 2011;40(2):82-97. [doi: 10.1080/16506073.2011.573807] [Medline: 25155812]
39. Rozental A, Andersson G, Boettcher J, Ebert DD, Cuijpers P, Knaevelsrud C, et al. Consensus statement on defining and measuring negative effects of internet interventions. Internet Interventions 2014 Mar;1(1):12-19. [doi: 10.1016/j.invent.2014.02.001]
40. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
41. Nowell LS, Norris JM, White DE, Moules NJ. Thematic analysis: striving to meet the trustworthiness criteria. Int J Qual Methods 2017 Oct 02;16(1). [doi: 10.1177/1609406917733847]
42. Chan JK, Farrer LM, Gulliver A, Bennett K, Griffiths KM. University students' views on the perceived benefits and drawbacks of seeking help for mental health problems on the internet: a qualitative study. JMIR Hum Factors 2016 Jan 19;3(1):e3 [FREE Full text] [doi: 10.2196/humanfactors.4765] [Medline: 27026140]
43. Carter J, Sandall J, Shennan AH, Tribe RM. Mobile phone apps for clinical decision support in pregnancy: a scoping review. BMC Med Inform Decis Mak 2019 Nov 12;19(1):219 [FREE Full text] [doi: 10.1186/s12911-019-0954-1] [Medline: 31718627]
44. Battle JD, Farrow L, Tibaijuka J, Mitchell M. mHealth for Safer Deliveries: a mixed methods evaluation of the effect of an integrated mobile health intervention on maternal care utilization. Healthc (Amst) 2015 Dec;3(4):180-184. [doi: 10.1016/j.hjdsi.2015.10.011] [Medline: 26699340]
45. Marko KI, Krapf JM, Meltzer AC, Oh J, Ganju N, Martinez AG, et al. Testing the feasibility of remote patient monitoring in prenatal care using a mobile app and connected devices: a prospective observational trial. JMIR Res Protoc 2016 Nov 18;5(4):e200 [FREE Full text] [doi: 10.2196/resprot.6167] [Medline: 27864167]
46. Donker T, Petrie K, Proudfoot J, Clarke J, Birch MR, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013 Nov 15;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]
47. O'Mahen HA, Flynn HA. Preferences and perceived barriers to treatment for depression during the perinatal period. J Womens Health (Larchmt) 2008 Oct;17(8):1301-1309. [doi: 10.1089/jwh.2007.0631] [Medline: 18816202]
48. Woolhouse H, Brown S, Krastev A, Perlen S, Gunn J. Seeking help for anxiety and depression after childbirth: results of the Maternal Health Study. Arch Womens Ment Health 2009 Apr 13;12(2):75-83. [doi: 10.1007/s00737-009-0049-6] [Medline: 19214705]
49. Dennis C, Chung-Lee L. Postpartum depression help-seeking barriers and maternal treatment preferences: a qualitative systematic review. Birth 2006 Dec;33(4):323-331. [doi: 10.1111/j.1523-536X.2006.00130.x] [Medline: 17150072]

50. Barak A, Hen L, Boniel-Nissim M, Shapira N. A comprehensive review and a meta-analysis of the effectiveness of internet-based psychotherapeutic interventions. J Technol Human Services 2008 Jul 03;26(2-4):109-160. [doi: 10.1080/15228830802094429]
51. Barak A, Klein B, Proudfoot JG. Defining internet-supported therapeutic interventions. Ann Behav Med 2009 Aug;38(1):4-17. [doi: 10.1007/s12160-009-9130-7] [Medline: 19787305]
52. Farrer L, Christensen H, Griffiths KM, Mackinnon A. Web-based cognitive behavior therapy for depression with and without telephone tracking in a national helpline: secondary outcomes from a randomized controlled trial. J Med Internet Res 2012 Jun 27;14(3):e68 [FREE Full text] [doi: 10.2196/jmir.1859] [Medline: 22738715]
53. Fonseca A, Gorayeb R, Canavarro MC. Women's use of online resources and acceptance of e-mental health tools during the perinatal period. Int J Med Inform 2016 Oct;94:228-236. [doi: 10.1016/j.ijmedinf.2016.07.016] [Medline: 27573331]
54. Preschl B, Wagner B, Forstmeier S, Maercker A. E-health interventions for depression, anxiety disorders, dementia, and other disorders in old age: a review. J Cyber Ther Rehabil 2011:4.
55. Kingston DE, Biringer A, McDonald SW, Heaman MI, Lasiuk GC, Hegadoren KM, et al. Preferences for mental health screening among pregnant women: a cross-sectional study. Am J Prev Med 2015 Oct;49(4):e35-e43. [doi: 10.1016/j.amepre.2015.03.026] [Medline: 26143952]
56. Kingston D, Austin M, Heaman M, McDonald S, Lasiuk G, Sword W, et al. Barriers and facilitators of mental health screening in pregnancy. J Affect Disord 2015 Nov 01;186:350-357. [doi: 10.1016/j.jad.2015.06.029] [Medline: 26281038]
57. Reay R, Matthey S, Ellwood D, Scott M. Long-term outcomes of participants in a perinatal depression early detection program. J Affect Disord 2011 Mar;129(1-3):94-103. [doi: 10.1016/j.jad.2010.07.035] [Medline: 20800898]
58. Goodman JW. Women's attitudes, preferences, and perceived barriers to treatment for perinatal depression. Birth 2009 Mar;36(1):60-69. [doi: 10.1111/j.1523-536X.2008.00296.x] [Medline: 19278385]
59. Batterham PJ, Calear AL. Preferences for internet-based mental health interventions in an adult online sample: findings from an online community survey. JMIR Ment Health 2017 Jun 30;4(2):e26. [doi: 10.2196/mental.7722]
60. Wallin EE, Mattsson S, Olsson EM. The preference for internet-based psychological interventions by individuals without past or current use of mental health treatment delivered online: a survey study with mixed-methods analysis. JMIR Ment Health 2016 Jun 14;3(2):e25 [FREE Full text] [doi: 10.2196/mental.5324] [Medline: 27302200]
61. Klein B, Cook S. Preferences for e-mental health services amongst an online Australian sample? E J Applied Psychol 2010 May 25;6(1):28-39. [doi: 10.7790/ejap.v6i1.184]
62. Casey L, Wright MA, Clough BA. Comparison of perceived barriers and treatment preferences associated with internet-based and face-to-face psychological treatment of depression. Int J Cyber Behav Psychol Learn 2014;4(4):16-22. [doi: 10.4018/ijcbpl.2014100102]
63. Horgan A, Sweeney J. Young students' use of the internet for mental health information and support. J Psychiatr Ment Health Nurs 2010 Mar;17(2):117-123. [doi: 10.1111/j.1365-2850.2009.01497.x] [Medline: 20465756]
64. Mohr DC, Siddique J, Ho J, Duffecy J, Jin L, Fokuo JK. Interest in behavioral and psychological treatments delivered face-to-face, by telephone, and by internet. Ann Behav Med 2010 Aug;40(1):89-98 [FREE Full text] [doi: 10.1007/s12160-010-9203-7] [Medline: 20652466]
65. Rizvi SL, Dimeff LA, Skutch J, Carroll D, Linehan MM. A pilot study of the DBT coach: an interactive mobile phone application for individuals with borderline personality disorder and substance use disorder. Behav Ther 2011 Dec;42(4):589-600. [doi: 10.1016/j.beth.2011.01.003] [Medline: 22035988]
66. Watts S, Mackenzie A, Thomas C, Griskaitis A, Mewton L, Williams A, et al. CBT for depression: a pilot RCT comparing mobile phone vs. computer. BMC Psychiatry 2013 Feb 07;13:49 [FREE Full text] [doi: 10.1186/1471-244X-13-49] [Medline: 23391304]
67. Kenny R, Dooley B, Fitzgerald A. Feasibility of "CopeSmart": a telemental health app for adolescents. JMIR Ment Health 2015 Aug 10;2(3):e22 [FREE Full text] [doi: 10.2196/mental.4370] [Medline: 26552425]
68. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]

Abbreviations

HOPE: Healthy Outcomes of Pregnancy and Postpartum Experiences
IPT: interpersonal psychotherapy

Edited by A Mavragani; submitted 19.02.22; peer-reviewed by C Ringham, B Parker; comments to author 10.04.22; revised version received 30.04.22; accepted 01.05.22; published 10.06.22.

Please cite as:
Bright KS, Stuart S, Mcneil DA, Murray L, Kingston DE
Feasibility and Acceptability of Internet-Based Interpersonal Psychotherapy for Stress, Anxiety, and Depression in Prenatal Women: Thematic Analysis
JMIR Form Res 2022;6(6):e23879
URL: https://formative.jmir.org/2022/6/e23879
doi:10.2196/23879
PMID:35687403

Original Paper

A Theory-Informed, Personalized mHealth Intervention for Adolescents (Mobile App for Physical Activity): Development and Pilot Study

Alex Domin[1], BA, MPhil, PhD; Arif Uslu[1], BSc, MSc; André Schulz[1], BSc, MSc, PhD; Yacine Ouzzahra[2], BSc, MSc, PhD; Claus Vögele[1], BSc, MSc, PhD

[1]Research Group: Self-Regulation and Health, Department of Behavioural and Cognitive Sciences, University of Luxembourg, Esch-sur-Alzette, Luxembourg

[2]Research Support Department, University of Luxembourg, Esch-sur-Alzette, Luxembourg

Corresponding Author:
Alex Domin, BA, MPhil, PhD
Research Group: Self-Regulation and Health
Department of Behavioural and Cognitive Sciences
University of Luxembourg
Maison des Sciences Humaines
11, Porte des Sciences
Esch-sur-Alzette, 4366
Luxembourg
Phone: 352 46 66 44 93
Fax: 352 46 66 44 39389
Email: alex.domin@uni.lu

Abstract

Background: Evidence suggests that physical activity (PA) during childhood and adolescence is crucial as it usually results in adequate PA levels in adulthood. Given the ubiquitous use of smartphones by adolescents, these devices may offer feasible means to reach young populations and deliver interventions aiming to increase PA participation and decrease sedentary time. To date, very few studies have reported smartphone-based interventions promoting PA for adolescents. In addition, most available fitness apps do not include the latest evidence-based content.

Objective: This paper described the systematic development of a behavior change, theory-informed Mobile App for Physical Activity intervention with personalized prompts for adolescents aged 16 to 18 years. The within-subject trial results provided the first evidence of the general effectiveness of the intervention based on the outcomes step count, sedentary time, and moderate to vigorous PA (MVPA) minutes. The effectiveness of the intervention component *personalized PA prompt* was also assessed.

Methods: A 4-week within-subject trial with 18 healthy adolescents aged 16 to 18 years was conducted (mean age 16.33, SD 0.57 years). After the baseline week, the participants used the Mobile App for Physical Activity intervention (Fitbit fitness tracker+app), which included a daily personalized PA prompt delivered via a pop-up notification. A paired 1-tailed *t* test was performed to assess the effectiveness of the intervention. Change-point analysis was performed to assess the effectiveness of a personalized PA prompt 30 and 60 minutes after prompt delivery.

Results: The results showed that the intervention significantly reduced sedentary time in adolescents during the first week of the trial (t_{17}=−1.79; *P*=.04; bootstrapped *P*=.02). This trend, although remaining positive, diminished over time. Our findings indicate that the intervention had no effect on metabolic equivalent of task–based MVPA minutes, although the descriptive increase may give reason for further investigation. Although the results suggested no overall change in heart rate–based MVPA minutes, the results from the change-point analyses suggest that the personalized PA prompts significantly increased heart rate per minute during the second week of the study (t_{16}=1.84; *P*=.04; bootstrapped *P*=.04). There were no significant increases in participants' overall step count; however, the personalized PA prompts resulted in a marginally significant increase in step counts per minute in the second week of the study (t_{17}=1.35; *P*=.09; bootstrapped *P*=.05).

Conclusions: The results of the trial provide preliminary evidence of the benefit of the Mobile App for Physical Activity intervention for modest yet significant reductions in participants' sedentary time and the beneficial role of personalized PA

prompts. These results also provide further evidence of the benefits and relative efficacy of personalized activity suggestions for inclusion in smartphone-based PA interventions. This study provides an example of how to guide the development of smartphone-based mobile health PA interventions for adolescents.

(JMIR Form Res 2022;6(6):e35118) doi:10.2196/35118

KEYWORDS

mobile health; physical activity; app; adolescents; within-subject; mHealth; sedentary behavior; behavior change techniques; BCTs; Fitbit; mobile phone

Introduction

Background

The beneficial impact of physical activity (PA) has been extensively documented, showing improved physical and mental health across the life span together with increased life expectancy [1]. In contrast, lack of PA and increased sedentary time continue to represent a serious public health burden. Low PA levels are associated with a higher prevalence of chronic diseases (eg, coronary heart disease, type 2 diabetes, breast cancer, and colon cancer) and an increased risk of morbidity and mortality, accounting for >5.3 million premature deaths annually [2]. In European Union countries, alarming levels of physical inactivity have been observed in adolescents [1]. As indicated by the Organisation for Economic Co-operation and Development report *Health at a Glance: Europe 2020*, less than 25% of boys and 20% of girls show sufficient levels of self-reported PA at the age of 15 years [3].

There is evidence that PA during childhood and adolescence is crucial as it usually results in adequate PA levels in adulthood [1]. Therefore, it is an essential research and public health priority to increase PA participation and decrease sedentary time in adolescents. Specific attention should be devoted to adolescents aged 16 to 18 years as they show the lowest absolute PA levels among children and young people aged 5 to 19 years [4]. Recent advances in digital technology have the potential to be successfully used in interventions aiming to improve PA-related outcomes in adolescents. In this study, we used the rationale that mobile devices have the potential to help their users engage in and adhere to different types of PA in several contexts (eg, school, leisure time, and transportation). In addition, mobile health (mHealth) interventions have the potential to adaptively respond to individuals' actions and states and deliver intervention options *just in time* (ie, when and where they are most appropriate [5,6]). Such just-in-time adaptive interventions (JITAIs) can facilitate health behavior change at times of both need for behavior support and receptivity [7]. Given the ubiquitous use of smartphones by adolescents, these devices may offer feasible means to reach young populations and deliver interventions aiming to increase PA participation and decrease sedentary time. Previous research suggests that an automated advice system may be as productive as or even preferable to a human advisor for increasing PA participation [8,9]. Preliminary evidence suggests that cardiorespiratory fitness gains can be sustained using a dedicated smartphone app [10-12]. The literature on JITAIs in particular shows mixed yet promising evidence. Specifically, a review found mixed evidence for JITAI effects on behavior, but no study was sufficiently powered to detect any effects. Another study reported that the JITAI condition demonstrated a significant improvement in health over the wait-list control condition [13,14]. In addition to interventions based on smartphone apps potentially being efficacious, they may also come with reduced costs in comparison with interventions involving face-to-face interviewing and guidance.

Although numerous fitness apps for smartphones address PA participation and sedentary behavior, most of them do not include the latest evidence-based content [15]. A recent review concluded that, despite not being grounded in theory, some interventions contain one or more behavior change components or behavior change techniques (BCTs) [16]. However, the systematic implementation of BCTs in dedicated apps has rarely been achieved. Michie et al [17] argue that many interventions applying recommended BCTs are not designed systematically and are theory-inspired rather than theory-based. This may result in low participant engagement with the intervention and a lack of longitudinal effects. Therefore, it is crucial to implement BCTs systematically while evaluating not only the effectiveness of the complete intervention but also the effectiveness of its smallest components. In a recent scoping review, we argued that the efficacy of smartphone-based mHealth PA interventions can be considerably improved through a more systematic approach of developing, reporting, and coding the interventions [18]. Therefore, in this study, we try to build on the previous theoretical findings systematically to maximize the impact of our intervention.

Smartphone-based behavior change interventions are typically tested using randomized controlled trials (RCTs) [19]. Although RCTs provide the highest level of scientific evidence, they evaluate the effect of an intervention as a whole. Therefore, RCTs would typically not provide information regarding which components work best and what factors modulate their efficacy. Considering the critical role of factors such as the timing of administration and the context in which components are implemented [6], an RCT may not provide the level of detail required to appropriately assess the efficacy of smartphone-based interventions. Finally, RCTs are quite long, averaging 5.5 years from recruitment to publication date [20,21]. Therefore, it was suggested to implement alternative designs that may provide prompter and more relevant answers while being more suitable to the research question. Trial designs where participants serve as their own controls, also known as within-subject designs, can reduce the number of study participants needed to detect outcomes and accelerate the research process while also simplifying the study procedures [22]. In addition, such designs have a much shorter duration

and allow for the testing of the efficacy of individual components [6]. Therefore, an mHealth PA intervention for adolescents may benefit from the implementation of such an evaluation design.

Personalization of intervention components has been shown to be important for their overall success as it may significantly affect the to date unresolved challenge of mHealth intervention engagement of the participants and, as a result, the outcomes of the intervention [23]. However, most current mHealth apps do not often offer personalized features although they could be important in increasing motivation and engagement. Examples of personalization include but are not limited to differentiation between habitual and unforeseen behaviors, collection of information about preferred PA, and activity suggestions depending on the current location or daily schedule [23,24]. Therefore, to maximize the potential of the intervention, it is important that the intervention components are personalized for its participants.

Previous studies mainly centered on the adult population have demonstrated that tackling the aforementioned gaps might be beneficial. Bond et al [25] developed and tested a smartphone-based intervention with a dedicated smartphone app, which significantly reduced sedentary behavior time over 4 weeks. Rabbi et al [23] and Klasnja et al [24] designed smartphone-based interventions that demonstrated preliminary evidence of the efficacy of personalized PA suggestions that are contextualized to the user's previous behavior and environment. Kramer et al [26] conducted a microrandomized trial (MRT) reporting a significant step goal increase triggered by cash incentive components. Finally, Gaudet et al [27] used a minimalist PA Fitbit tracker–based intervention with adolescents aged 13 to 14 years, which resulted in increased PA.

Objectives

This project aimed to further the findings from previously conducted studies while concentrating on the adolescent population. The Mobile App for Physical Activity intervention uses a personalization feature through setting individualized PA goals and delivering tailored feedback based on the individual's performance not limited solely to step count. The Mobile App for Physical Activity intervention was developed based on the Behaviour Change Wheel framework [28] by applying an approach that incorporates findings from qualitative studies and recommended and efficacious BCTs based on systematic reviews and meta-analyses. Finally, the research available to date tends to focus on the development of apps for adults rather than adolescents. Although adolescents aged 16 to 18 years represent a highly relevant target group for improving PA participation, only very few studies have addressed this particular age group [18]. To the best of our knowledge, this is the first study to develop a behavior change PA mHealth intervention with personalized prompts for adolescents aged 16 to 18 years evaluated using a within-subject experimental design.

The Mobile App for Physical Activity within-subject trial was conducted with the objective to provide the first evidence of the general effectiveness of the Mobile App for Physical Activity intervention among adolescents based on the outcomes of step count, sedentary time, and moderate to vigorous PA (MVPA) minutes. The effectiveness of the intervention component *personalized PA prompt* was also assessed using change-point analyses to determine whether similar PA smartphone-based interventions could benefit from the implementation of such a component.

We hypothesized that (1) the Mobile App for Physical Activity intervention would decrease the daily sedentary time of adolescents during the intervention in weeks 1, 2, and 3 compared with baseline measurements (primary variable of interest). We also hypothesized that (2) there would be an increase in the time spent in MVPA minutes and the number of daily steps in weeks 1, 2, and 3 compared with baseline measurements (secondary variables of interest). Finally, we hypothesized that (3) the participants' step count and heart rate (HR) would show an increase after the delivery of a personalized PA prompt.

Methods

The Mobile App for Physical Activity Intervention

Mobile App for Physical Activity is an intervention developed to promote PA among adolescents aged 16 to 18 years. It was aimed at adolescents who showed insufficient PA levels according to the World Health Organization (WHO) recommendations (ie, <60 minutes of moderate or vigorous PA each day, associated with 11,700 steps daily [29]) but were interested in increasing it. This minimalist, multicomponent mHealth PA intervention combines a Fitbit smartphone app [30], personalized assistance, and a wrist-worn activity tracker (Fitbit Charge 4) [31] that collects HR-based Active Zone Minutes or MVPA minutes, active minutes based on the metabolic equivalent of task (MET) or MVPA minutes, step count, and sedentary minutes based on MET data. The Mobile App for Physical Activity intervention includes basic features of the Fitbit app and tracker as well as additional personalized assistance features—daily personalized PA prompts, weekly goal adjustment, and interactive assistance realized via the chat messaging feature—to help participants resolve any problems concerning achieving daily goals or using the intervention components. All additional components were implemented just in time by AD using the back-end features of the intervention (the Fitabase platform and Fitbit web interface). In total, 4 different outcome measures were controlled for following the recommendations of Thompson et al [32] suggesting the consideration of a multidimensional PA user profile. This device was selected for several reasons. We were interested in a device that would be, on the one hand, attractive yet unnoticeable and the least burdensome (to support user engagement) yet, by contrast, able to collect HR data (for MVPA minute calculation based on HR data in addition to MET values) and that would track and automatically recognize various types of PA performed by users. The triaxial accelerometer produced by ActiGraph is considered the gold standard in PA measurement, currently proposing several wrist-worn activity trackers [33]. However, the proposed HR measurement method would have necessitated additional wireless devices, which may be considered

burdensome by users. In addition, activity recognition was not automated. The latest reviews have reported satisfactory validity, reliability, and feasibility of consumer-grade activity trackers produced by Fitbit and other companies [34,35]. After analyzing the market of commercial activity trackers, we identified the device that matched most of our requirements—Fitbit Charge 4, a small, wrist-worn, waterproof activity tracker with an inbuilt photoplethysmographic sensor to assess HR and automatic recognition of 7 types of PA. Similar Fitbit devices have already been used in mHealth PA interventions for data collection [27,36-40]. The limitation that we encountered with this device related to data received from the Fitbit server—MVPA minutes or, according to Fitbit, *activity minutes* (calculated based on MET values) and *Active Zone Minutes* (calculated based on HR values) were provided *as is* (ie, without providing an algorithm based on which MVPA minutes were calculated). To date, the literature presents mixed evidence on the validity of HR (–3% to +3% error rates), maximal aerobic capacity (VO_{2max}), and energy expenditure measurements [41] yet confirms the relative validity of the Fitbit MVPA minute calculation [42], accurate recognition of the PA type [43], excellent interinstrument reliability, and good levels of agreement between devices [44]. On the basis of these considerations, we decided to use the Fitbit Charge 4 as a primary data collection device taking into account the limitations described.

The Fitbit app presents the user with a large set of features and tools that can be displayed in a selective manner. To use this software as a part of the Mobile App for Physical Activity intervention, the functionality of the Fitbit app was used as a toolkit and was tweaked to only make use of the features selected below. Therefore, the next constituent of the Mobile App for Physical Activity intervention included a combination of the Fitbit smartphone app and personalized assistance, which included both *push* and *pull* intervention components.

The first pull component was *graphs and stats*. This component, delivered via the home page of the Fitbit app, presents users with graphical feedback on their daily goals with the potential for personalization. Specifically, users can compare the results attained thus far with their personalized goals. Each participant was assigned 2 personalized daily goals: a step count goal and an *Active Zone Minutes* goal. These goals were set individually via remote Fitbit account access at the end of every week based on a 5% increase from the average daily step or Active Zone Minute count achieved during the previous week. A 5% increase mark was chosen to provide a substantial yet feasible increase goal based on the findings from the study by Degroote et al [45]. If a participant underperformed, the daily step or Active Zone Minute goals remained the same.

The second pull component was *interactive assistance*. Through a dedicated message tab, the participants received advice or could communicate with AD in case they encountered any problems in either achieving their daily goals or using the intervention components. Web-based support was initially intended as an automated advisor providing personalized support and problem-solving strategies and answering inquiries based on a strategy grounded in artificial intelligence research. However, because of the prolonged development period, for this first version of the Mobile App for Physical Activity intervention, AD substituted an automated advice system. There is recent evidence that, while providing solutions for identified obstacles, the implementation of an automated advisor is considered promising in increasing PA even compared with human advisors. Therefore, this component should be further developed and tested in future trials [9,10,46,47]. This component also has the potential to reinforce coping planning (by providing coping responses for dealing with potential barriers and difficult situations) [26,48].

The first push component was a *personalized PA prompt*. According to the results of the 2 most recent MRTs, tailored push suggestions in a PA context were associated with greater engagement with an mHealth app and increased PA participation [24,49]. In the Mobile App for Physical Activity, the users received 1 tailored suggestion (delivered as a pop-up notification). Every day after school classes had finished (5 PM-7 PM), the participants received a personalized PA prompt via a Fitbit pop-up notification. Each message was prepared by AD considering the daily step count performance so far and its percentage correlation with the daily goal. Depending on the participant's achievements, the message could be framed in 5 different ways according to the percentage reached compared with the personalized goal: <40%, ≥40%, ≥60%, ≥80%, and ≥100%. The message also included different PA health benefits that the participants could potentially achieve by following their activity goals based on WHO recommendations [50]. Additional attention was paid to the positive framing of the PA suggestions [47]. The Mobile App for Physical Activity intervention included >25 PA suggestion templates developed by the research team using data collected during focus group discussions. All generated suggestion templates were reviewed and edited, and additional suggestions were created to provide a sufficient number of prompts for each condition, including the personalization to such events as an exam period. An example of such a suggestion would be the following: *Hi, Bob! You did around xxxxx steps so far and reached 80% of your goal for today—good job, can you do even better? Interesting fact: physical activity benefits improved concentration, so if you are active, you may get better study results! Keep up the effort!* Therefore, depending on the participants and their performance level as well as their daily goal, messages could differ in relation to five variables: (1) participant name, (2) number of steps so far, (3) step count percentage correlation with the daily goal, (4) PA health benefits based on WHO recommendations, and (5) general message framing. This resulted in a personalized message that could potentially be followed by interactive assistance, which is a qualitatively different combination compared with the generic messages currently provided by most commercial apps. In further iterations of the Mobile App for Physical Activity intervention, this process is planned to be more automated and personalized considering contextual factors.

The second push component was *reminders to move*. Every day after classes (4 PM-9 PM), if the participants' behavior was identified as sedentary (no steps or any other HR-increasing activity were performed), they would receive a pop-up reminder motivating them to take ≥250 steps by the end of each hour.

These approaches were designed to support the users' self-regulation, which is recognized as a principal factor in health behavior change [51,52]. The self-monitoring and feedback strategies implemented in this intervention component are characterized as "especially helpful" and recommended for inclusion in PA promotion interventions [47].

The last intervention component was *rewards*. This component was introduced to reinforce the self-monitoring strategies used in the Mobile App for Physical Activity intervention. The strategy of providing users with rewards was recently reported to be especially helpful in increasing PA [47]. In the preliminary-held focus group discussions, adolescents identified the sports-related rewards as the most attractive. At intake, the participants were informed that rewards could be obtained based on the number of consecutively accumulated days in line with or above the personalized goals. The rewards consisted of a digital gift voucher from a local sporting goods store. The value of the voucher depended on the number of consecutively accumulated days in line with or above the personalized goals (ie, 3, 5, or 7 days of the week resulted in a €15 [US $15.97], €30 [US $31.94], or €40 [US $42.58] voucher reward, respectively). Rewards or gift vouchers were time-contingent (ie, they were delivered via email directly at the end of the week). The reward scenario aimed to reinforce users' self-monitoring and the consequent regular performance of sufficient PA. As the Mobile App for Physical Activity intervention was intended to be available to users over a longer period (ie, after the completion of the study), incentives were projected to be effective at early stages and further gradually substituted by habitually formed self-monitoring strategies [53]. Finally, the participants were allowed to keep the activity trackers after the end of the study.

Selection of the Intervention Components

The selection of the Mobile App for Physical Activity intervention components was guided by 2 combined approaches. The first approach included an analysis of qualitative studies (survey, interview, and focus group discussion as methods of data collection) that explored user preferences in terms of the technological functionality of PA promotion apps [54-61]. We also conducted a focus group discussion to explore the app feature preferences of adolescents. After identification of potentially advantageous and attractive features for the PA promotion app, we aligned the results with our second approach (ie, the identification and implementation of recommended, effective, and efficacious BCTs).

To identify and select such BCTs, we used the approach described by Lyons et al [62], compiling recommended BCTs from several sources in 1 list [62]. Our selection was based on the following steps: (1) successful BCTs for increasing PA in adolescents based on the meta-analysis by Brannon and Cushing for adolescents [63], (2) BCTs that predicted PA as reported in meta-analyses on PA interventions for adults [51,63-67], (3) recommendations from the systematic review by Sullivan et al [47], and (4) BCTs identified by applying the Behaviour Change Wheel framework [28,68]. We used the Acceptability, Practicability, Effectiveness, Affordability, Side-effects, and Equity criteria (worksheet 7 of the Behaviour Change Wheel manual) to identify the appropriate BCTs based on the Mobile App for Physical Activity intervention functions. As an example, the BCT *Credible source* was identified as appropriate and applied within 2 intervention components: interactive assistance and personalized PA prompt. In the Mobile App for Physical Activity intervention, for both cases, a communication from a credible source in favor of the active behavior was presented. In a final step, the selection was reviewed and confirmed by a panel of 4 senior researchers (Textbox 1).

It is important to note the limitations of the current literature in which the presented BCTs were identified. First, the Mobile App for Physical Activity intervention was developed for adolescents, although data from meta-analyses and reviews for adults were used as there is a gap concerning studies on mHealth PA promotion in younger populations [18]. Second, although the review by Brannon and Cushing [63] concentrates on apps, the meta-analysis they performed was based on *classic* PA interventions rather than apps. Third, the identified meta-analyses (also based on *classic* PA interventions) presented a mix of results in terms of BCTs. Therefore, we included BCTs in our list only if they were associated with PA in at least two of the 6 meta-analyses. Finally, yet importantly, there are several reviews on the topic published to date that can inform the reader on BCTs identified in efficacious interventions [11,69,70]; however, only 1 review (Sullivan et al [47]) provides specific recommendations on helpful strategies [47]. Therefore, we implemented the recommendations from these reviews and meta-analyses in that we coded all the BCTs presented according to previous versions of BCT taxonomies (26 and 40 BCTs) into the latest taxonomy (93 BCTs) [71-73].

The combination of potentially advantageous and attractive PA app features and the selection of BCTs associated with PA change supported a novel approach in developing the Mobile App for Physical Activity intervention. The selection of BCT components was based on two factors: (1) components were selected if considered attractive to adolescents and (2) components were selected if they were considered the most appropriate to reflect a certain BCT (eg, components such as *graphical representation of performed PA* would naturally include the BCT *Feedback on behavior*). The result (ie, the components of the app and the BCTs implemented in them) is presented in Table 1.

Textbox 1. Recommended, effective, and efficacious behavior change techniques (BCTs).

BCTs associated with physical activity (PA) identified in at least two of the 6 meta-analyses [51,63-67]

- Self-monitoring of behaviour (2.3), self-monitoring of outcome(s) of behaviour (2.4)
- Goal setting (behaviour; 1.1), goal setting (outcome; 1.3)
- Feedback on behaviour (2.2), feedback on outcome(s) of behaviour (2.7)
- Information about health consequences (5.1)

BCTs associated with PA according to the meta-analysis by Brannon and Cushing [63]

- Information about health consequences (5.1)
- Information about others' approval (6.3)
- Goal setting (behaviour; 1.1), goal setting (outcome; 1.3)
- Self-monitoring of behaviour (2.3), self-monitoring of outcome(s) of behaviour (2.4)
- Behavioural contract (1.8)

BCTs recommended for implementation by Sullivan et al [47]

- Goal setting: goal setting (behaviour; 1.1), goal setting (outcome; 1.3)
- Self-monitoring: self-monitoring of behaviour (2.3), self-monitoring of outcome(s) of behaviour (2.4)
- Feedback: feedback on behaviour (2.2), feedback on outcome(s) of behaviour (2.7)
- Rewards: reward (outcome; 10.10), nonspecific reward (10.3)
- Social support: social support (unspecified; 3.1), social support (practical; 3.2), social support (emotional; 3.3)
- Coaching: instruction on how to perform the behaviour (4.1)
- Identifying obstacles: problem solving (1.2)
- Restructuring negative attitudes: framing/reframing (13.2)
- Action planning: action planning (1.4)
- Modifying environmental factors: restructuring the physical environment (12.1)

BCTs selected through the Behaviour Change Wheel framework [28,68]

- Credible source (9.1)
- Monitoring of behaviour by others without feedback (2.1), monitoring of outcome(s) of behaviour without feedback (2.5)
- Adding objects to the environment (12.5)
- Review behaviour goal(s) (1.5), review outcome goal(s) (1.7)

Table 1. Behavior change techniques (BCTs) included in the Mobile App for Physical Activity intervention components.

Push or pull	Component	BCTs
Push	Personalized PA[a] prompt	Feedback on behavior, prompts/cues, discrepancy between current behavior and goal, information about health consequences, credible source
Push	Reminders to move	Feedback on behavior, prompts/cues, discrepancy between current behavior and goal
Push	Rewards	Reward (outcome), self-monitoring of behavior
Pull	Interactive assistance	Problem solving, restructuring the physical environment, credible source
Pull	Graphs and stats	Goal setting (outcome), feedback on behavior, information about health consequences, monitoring of outcome(s) of behavior without feedback

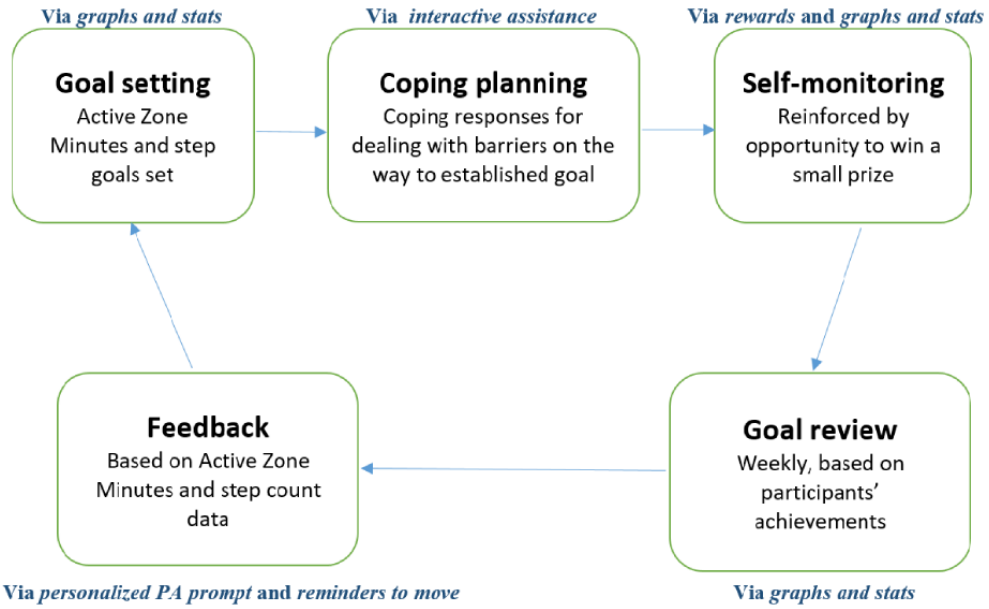
[a]PA: physical activity.

Theoretical Background

A meta-regression by Michie et al [51] has demonstrated that including such BCTs as self-monitoring combined with at least one other BCT from control theory (ie, prompt intention formation, prompt specific goal setting, providing feedback on performance, prompt self-monitoring of behavior, and prompt review of behavioral goals) in PA interventions is effective [51]. Three of these BCTs—namely, self-monitoring, goal setting, and feedback—correspond to the process of self-regulation or,

more specifically, control theory [74]. This theory suggests that self-monitoring behavior, receiving feedback, setting goals, and reviewing goals following feedback are central to behavioral self-management. Thus, the Mobile App for Physical Activity intervention components were centered on behavioral self-regulation (Figure 1). This, in addition to the aforementioned rigorous selection approach, enabled the selection and sequencing of the central BCTs [75].

Figure 1. Components of adjusted self-regulation control theory, which informed the Mobile App for Physical Activity intervention. PA: physical activity.



Participants

The selection criteria were similar to the ones applied in our focus group discussion study. A local international school was chosen to recruit healthy adolescents aged 16 to 18 years. The advertisement with study details was disseminated via the school's email service, seeking to attract adolescents who were insufficiently active yet, on principle, willing to increase their PA participation. In addition, several adolescent students of local Luxembourgish schools were recruited from the participant list of the focus group discussion, which took place earlier and has been described elsewhere [76]. The advertisement contained general information about the study and informed potential participants that they would be remunerated for taking part. Remuneration was provided by letting the participants keep the activity tracker implemented in the study (Fitbit Charge 4) and the possibility to win sporting goods store vouchers. The participants needed to be fluent in English and possess a smartphone. Participants were excluded if they had any constraints toward performing PA and if they owned and actively used an activity tracker (eg, Fitbit, Garmin, or Apple Watch) as additional devices provided within the Mobile App for Physical Activity trial could be perceived as burdensome. The study participants were provided with an informed consent form, which had to be signed by the participants themselves (when aged 18 years) or their legal representative (when aged <18 years) before participation.

Ethics Approval

This study was approved by the Ethics Review Panel of the University of Luxembourg (19-046A2 Mobile App for Physical Activity).

Study Design and Procedure

The Mobile App for Physical Activity intervention was carried out as a 4-week within-subject trial (baseline week+3 intervention weeks). At the selection stage, the PA profile of the interested participants was evaluated using the Physical Activity Questionnaire for Adolescents (PAQ-A) [77]. This questionnaire was selected based on an expert panel ranking that evaluated the PAQ-A as one of the very few self-report instruments with acceptable reliability, practicality, and validity [78]. The PAQ-A is a 7-day-recall self-administered questionnaire designed to provide a general estimate of PA levels in healthy adolescents aged between 8 and 20 years derived from a series of questionnaire items on activity during and after school, sports participation, and activity in the evenings and weekends [77]. Participants were included if they had a PAQ-A score of ≤3 out of 5 (low to moderate levels of weekly PA). An introduction session was organized for the participants at intake to provide them with general information; create Fitbit accounts; link the fitness trackers to them; and analyze their PA habits, basic sociodemographic information, and previous experiences with PA and fitness apps. The participants were then given activity trackers and instructed to wear them at all times (including bathing, sleeping, and swimming). The participants were also instructed to install the Fitbit app and create a Fitbit account while all the notifications for the Fitbit app (and activity tracker) were disabled during the baseline week, and users were encouraged to ignore any notifications that might appear erroneously. Credentials for the Fitbit account were shared with AD to further turn on certain notifications and features during the treatment weeks. The participants were also asked not to use the Fitbit app during the baseline week. AD also connected with all the participants via the Fitbit app (*Add*

friends feature) to send them later the personalized PA prompt via the dedicated *Messages* chat tab in the Fitbit app. After the initial week (a baseline period), notifications were turned on remotely via Fitbit account manipulation for proper functioning of components, and personalized step and Active Zone Minute goals were set. Users were informed that the Fitbit app could be assessed at any time without a prerequired frequency. Adherence to both the Fitbit app and the activity tracker and data collection was monitored through the Fitabase platform [79]. Emails were sent to the participants when tracker data were missing for >24 hours or when the participants forgot to synchronize data. If data were not received after 48 hours, the participant was contacted by the study supervisor via SMS text message or a phone call. Normally, this situation did not occur more often than 4 to 6 times per week.

Statistical Analysis

The aggregated data set was downloaded in CSV format from the Fitbit application programming interface via the Fitabase platform. Fitbit provides data of a different resolution (days or minutes) and, for the primary analysis, we aggregated daily into weekly records.

To test the first and second hypotheses, we [80] performed paired 1-tailed *t* tests to reveal within-subject differences in the primary and secondary outcome measures between measurement occasions (baseline, week 1, week 2, and week 3). We tested for the outcome measures outlined in Textbox 2.

Textbox 2. Primary and secondary outcome measures.

Primary outcome

- Sedentary minutes based on the metabolic equivalent of task (MET)

Secondary outcomes

- Active Minutes based on MET or moderate to vigorous physical activity (MVPA) minutes
- Active Zone Minutes based on heart rate or MVPA minutes
- Step count

As the 2 recruitment sources (international school and local schools) had different holiday schedules, we plotted for differences in outcomes between participants from these 2 groups. Owing to the small sample size, we computed a bootstrapped paired 1-tailed *t* test in addition to the classic paired 1-tailed *t* test. The measurement occasions were 7-day periods, further referred to as *baseline*, *week 1*, *week 2*, and *week 3*. Each week 1 to week 3 value was compared with the baseline. The α level was set to .05 for all tests. While α inflation is an issue with multiple testing, we decided not to use the Bonferroni correction as it is considered overly conservative and, therefore, increases the risk of type 2 error. Given the early stage of development and research in this area, we would argue that reporting the uncorrected results contributes to the literature by stimulating studies with larger samples in which hypotheses can then be tested more rigorously.

To test the third hypothesis, we performed change-point detection analysis to identify change points in the means of the step count and the HR time series in the minute-to-minute resolution. As personalized PA prompts were given every day between 5 PM and 7 PM, we used a time frame of 3 PM to 7 PM to estimate change points. Once estimated, we chose the best-fitting change point after 5 PM per participant and day. We then subtracted the averaged step counts and HR of the 30-minute period before from the 30-minute period after the change point to calculate the magnitude of physiological change at that time. The 30-minute period was chosen based on the approach by Klasnja et al [24]. As a 30-minute period value was rather theory-inspired, we used an empirical approach and tested for the magnitude of physiological change also within a 60-minute period. To analyze the differences between weeks, we aggregated the magnitude of HR changes per participant and week. Although we did not provide a personalized PA prompt during the baseline, we nevertheless controlled for the random variations from 5 PM to 7 PM, which resulted in a comparison value for the treatment weeks. All statistical analyses were carried out using RStudio (version 4.1.1; R Foundation for Statistical Computing) [80].

Results

Sample

We recruited 18 participants, of whom 6 (33%) were women. Most participants (11/18, 61%) were recruited from an international school, whereas the other 39% (7/18) were recruited from Luxembourgish schools. The participants had different levels of motivation to improve their PA behavior, varying from very low to moderate. For most of the participants (11/18, 61%)), English was their primary language. Nevertheless, plain English was used within the Mobile App for Physical Activity intervention to make it more attractive to all the participants. The participant mean age was 16.33 (SD 0.57) years. All participants (18/18, 100%) filled in the PAQ-A questionnaire, ranging from a score of 1, which indicates a low PA level, to a score of 5, which indicates a high PA level, with a mean score of 2.72 (SD 0.48). Most participants (16/18, 89%) owned an iOS-powered smartphone, with 11% (2/18) of the participants owning an Android-powered smartphone. We conducted the analysis while differentiating (color coding) between participants from the international school and Luxembourgish schools. However, the recruitment site was not introduced as a between-subject factor owing to the small sample size; therefore, we did not further report data on the differences between recruitment sites, describing rather our general observations of the differences between the 2 sites. Although

showing similar trends in sedentary behavior, the step count and MVPA minute trends were generally divergent; students from local schools tended to increase their step count and MVPA minutes, whereas students from the international school tended to decrease their step count and MVPA minutes over the trial period.

Primary Outcome Analysis: Change in Daily Time Spent in Sedentary Behavior Based on MET

Sedentary minute counts decreased significantly during the first week of the trial compared with the baseline (t_{17}=−1.79; *P*=.04; bootstrapped *P*=.02; Table 2). This effect diminished over time and was no longer significant at week 2 (t_{17}=−0.51; *P*=.30; bootstrapped *P*=.30) and week 3 (t_{17}=−0.94; *P*=.17; bootstrapped *P*=.21). Figure 2 depicts the time course of sedentary minutes over the entire duration of the intervention. The error area in this and the other figures represents the SE of the mean.

Table 2. Outcome measures during the baseline and intervention periods.

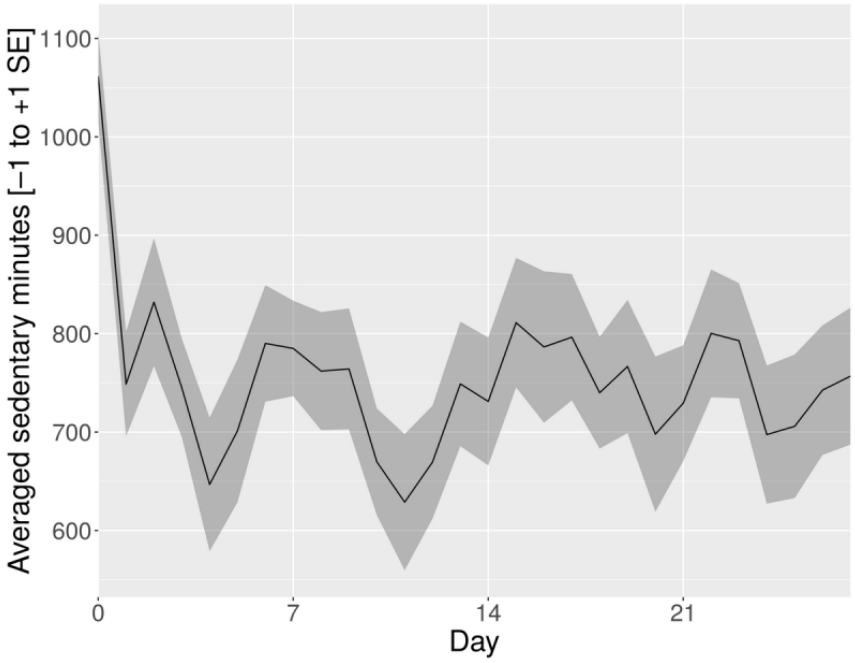
Outcome and measurement occasion	Values per day, mean (SD)
Sedentary behavior (minutes per week)	
Baseline	789.34 (176.71)
Week 1	718.30 (177.58)
Week 2	764.09 (209.89)
Week 3	749.13 (209.80)
MVPA [a] based on MET[b] (minutes per week)	
Baseline	90.37 (47.94)
Week 1	100.85 (46.24)
Week 2	95.50 (55.08)
Week 3	91.57 (49.10)
MVPA based on HR [c] (minutes per week)	
Baseline	42.88 (27.41)
Week 1	45.88 (27.15)
Week 2	37.81 (24.51)
Week 3	39.94 (32.66)
Step count (steps per week)	
Baseline	14,134.56 (4980.45)
Week 1	14,298.28 (4523.17)
Week 2	13,581.28 (5529.25)
Week 3	14,126.29 (5413.41)

[a]MVPA: moderate to vigorous physical activity.

[b]MET: metabolic equivalent of task.

[c]HR: heart rate.

Figure 2. Averaged sedentary minutes (mean+SE of the mean).

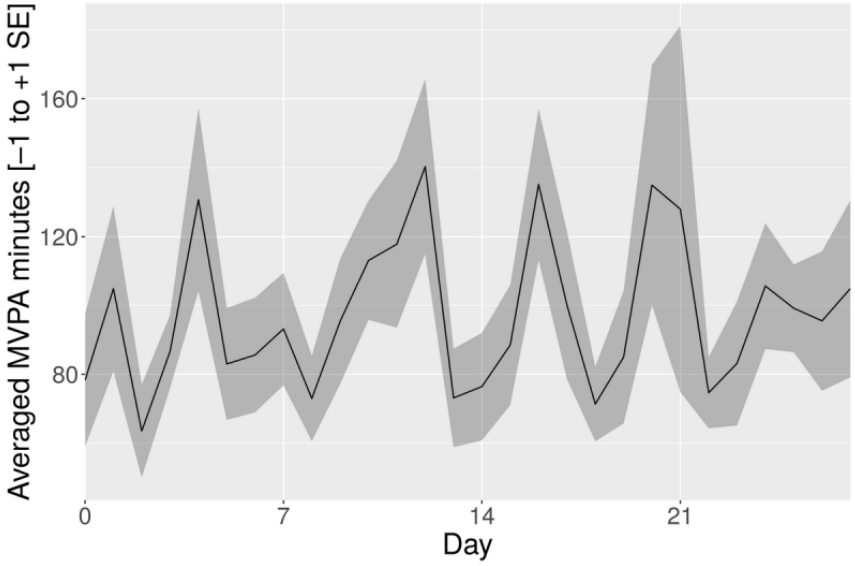


Secondary Outcome Analyses

Change in MVPA Minutes Based on MET

In total, 2 outliers with >500 MVPA minutes a day were excluded. With MET-based MVPA minutes, we observed a reversed nonsignificant trend in comparison with sedentary minute count (Table 2)—in the first week, the descriptive values of MET-based MVPA minutes increased compared with the baseline (t_{17}=1.23; *P*=.11; bootstrapped *P*=.12). This effect diminished over time and, although it was still positive in week 2 (t_{17}=0.41; *P*=.34; bootstrapped *P*=.34), it was smaller in week 3 (t_{17}=0.12; *P*=.45; bootstrapped *P*=.45). None of these changes reached significance levels (see Figure 3 for the complete time course).

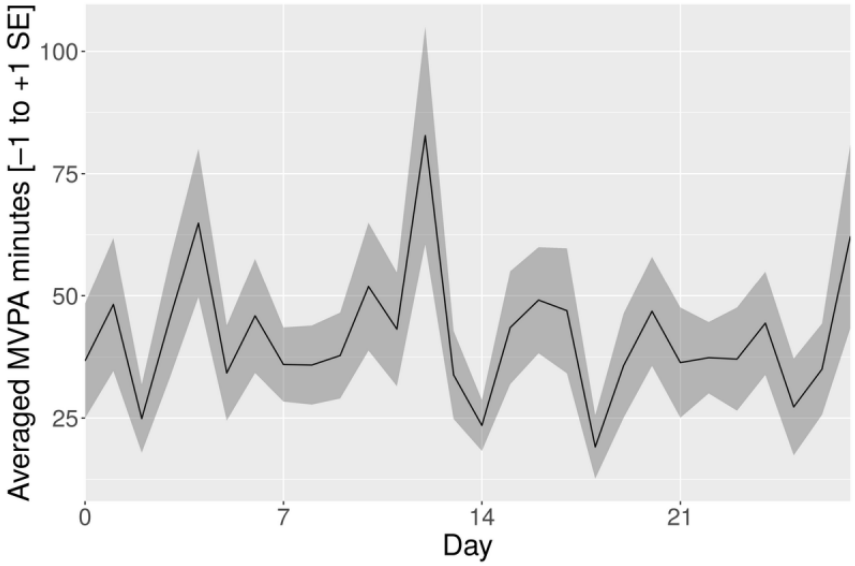
Figure 3. Averaged moderate to vigorous physical activity (MVPA) minutes based on the metabolic equivalent of task (mean+SE of the mean).



Change in MVPA Minutes Based on HR

Although we observed an initial nonsignificant increase in the first week (t_{17}=0.50; *P*=.31; bootstrapped *P*=.29), we observed a nonsignificant decline compared with the baseline for weeks 2 (t_{17}=−0.57; *P*=.71; bootstrapped *P*=.72) and 3 (t_{17}=−0.39; *P*=.65; bootstrapped *P*=.63). See Figure 4 for the complete time course.

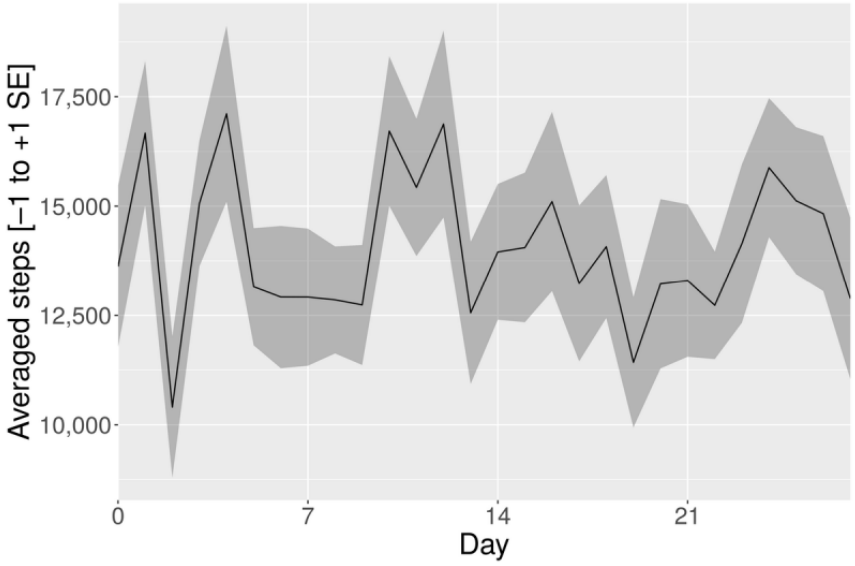
Figure 4. Averaged moderate to vigorous physical activity (MVPA) minutes based on heart rate (mean+SE of the mean).



Change in Step Count

The analysis of step count revealed a descriptive course similar to the results of the MVPA minutes based on HR (Table 2)—although we observed a slight nonsignificant increase in the first week (t_{17}=0.20; *P*=.41; bootstrapped *P*=.43), there was no clear linear trend over the following weeks. In weeks 2 (t_{17}=−0.33; *P*=.62; bootstrapped *P*=.63) and 3 (t_{17}=−0.006; *P*=.50; bootstrapped *P*=.50), we observed a nonsignificant decline compared with the baseline (Figure 5).

Figure 5. Averaged steps (mean+SE of the mean).



Estimation of the Effect After the Delivery of a Personalized PA Prompt: Change-Point Analyses

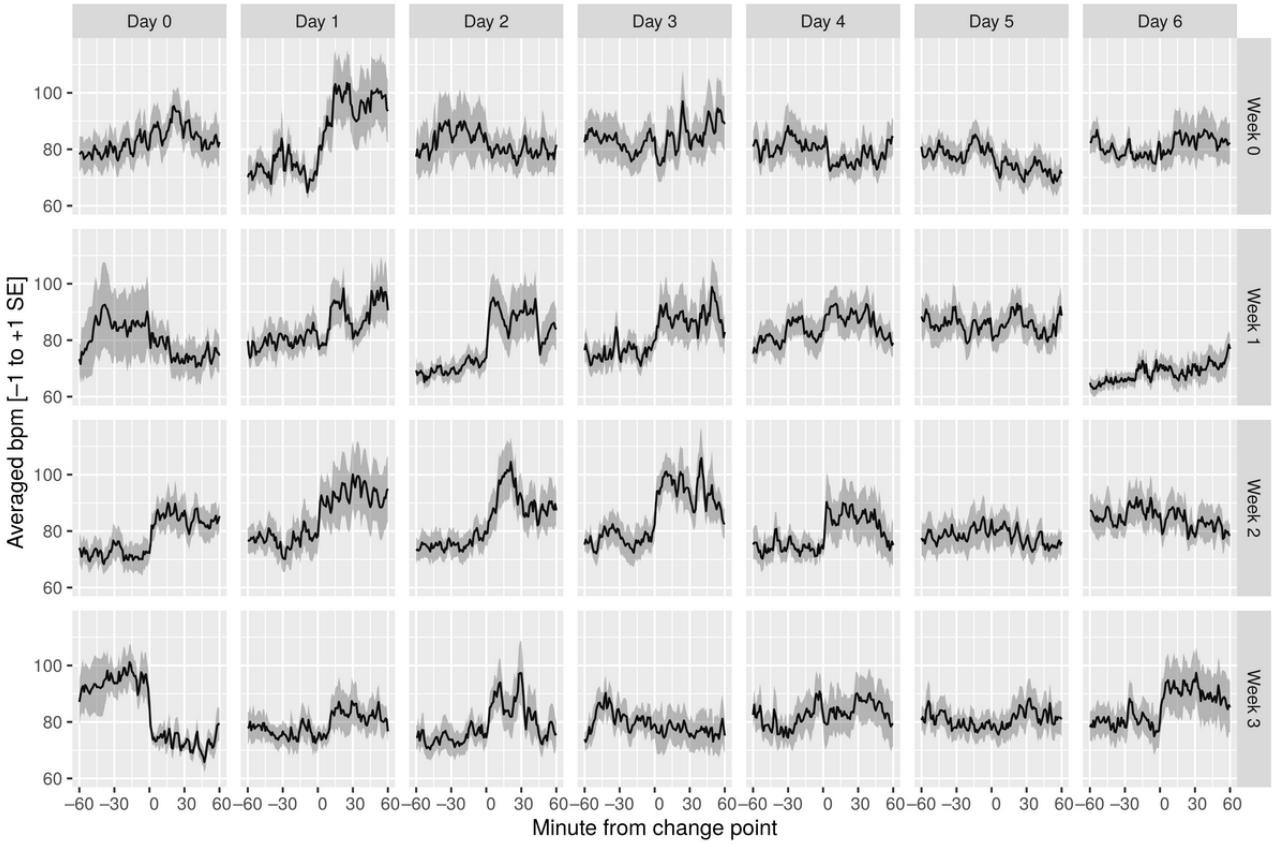
HR at 60 Minutes

We linearly interpolated some missing data (up to 5 minutes per day) to be able to perform the change-point analysis. If >5 minutes per day were missing, data from that day were excluded before running the change-point analysis. As shown in Table 3, the intervention resulted in a nonsignificant increase in HR compared with the baseline in the first week (t_{16}=1.28; *P*=.10; bootstrapped *P*=.09) and a significant increase in the second week (t_{16}=1.84; *P*=.04; bootstrapped *P*=.04). However, this effect diminished during week 3 (t_{16}=−0.07; *P*=.52; bootstrapped *P*=.52), where we observed a nonsignificant decline compared with the baseline (see Figure 6 for the complete time course).

Table 3. Change-point analysis results.

Outcome and measurement occasion	Minutes before	Minutes after	Increase per minute, mean (SD)
Change-point analysis: heart rate at 60 minutes (bpm)			
Baseline	80.637	81.762	1.12 (11.09)
Week 1	77.617	84.211	6.59 (11.73)
Week 2	78.877	89.024	10.14 (16.23)
Week 3	83.183	80.831	−2.35 (17.29)
Change-point analysis: heart rate at 30 minutes (bpm)			
Baseline	81.256	81.893	0.63 (13.28)
Week 1	78.656	84.534	5.87 (10.36)
Week 2	79.496	90.772	11.27 (15.88)
Week 3	83.647	81.865	−1.78 (15.74)
Change-point analysis: step count at 60 minutes (steps)			
Baseline	16.955	23.604	6.64 (13.10)
Week 1	19.813	25.510	5.69 (17.28)
Week 2	13.912	27.997	14.08 (21.92)
Week 3	17.826	23.755	5.92 (12.70)
Change-point analysis: step count at 30 minutes (steps)			
Baseline	19.732	27.038	7.30 (16.18)
Week 1	20.191	29.808	9.61 (16.50)
Week 2	15.036	30.855	15.81 (22.52)
Week 3	19.794	26.224	6.42 (13.33)

Figure 6. Change-point analysis: heart rate (mean+SE of the mean). bpm: beats per minute.



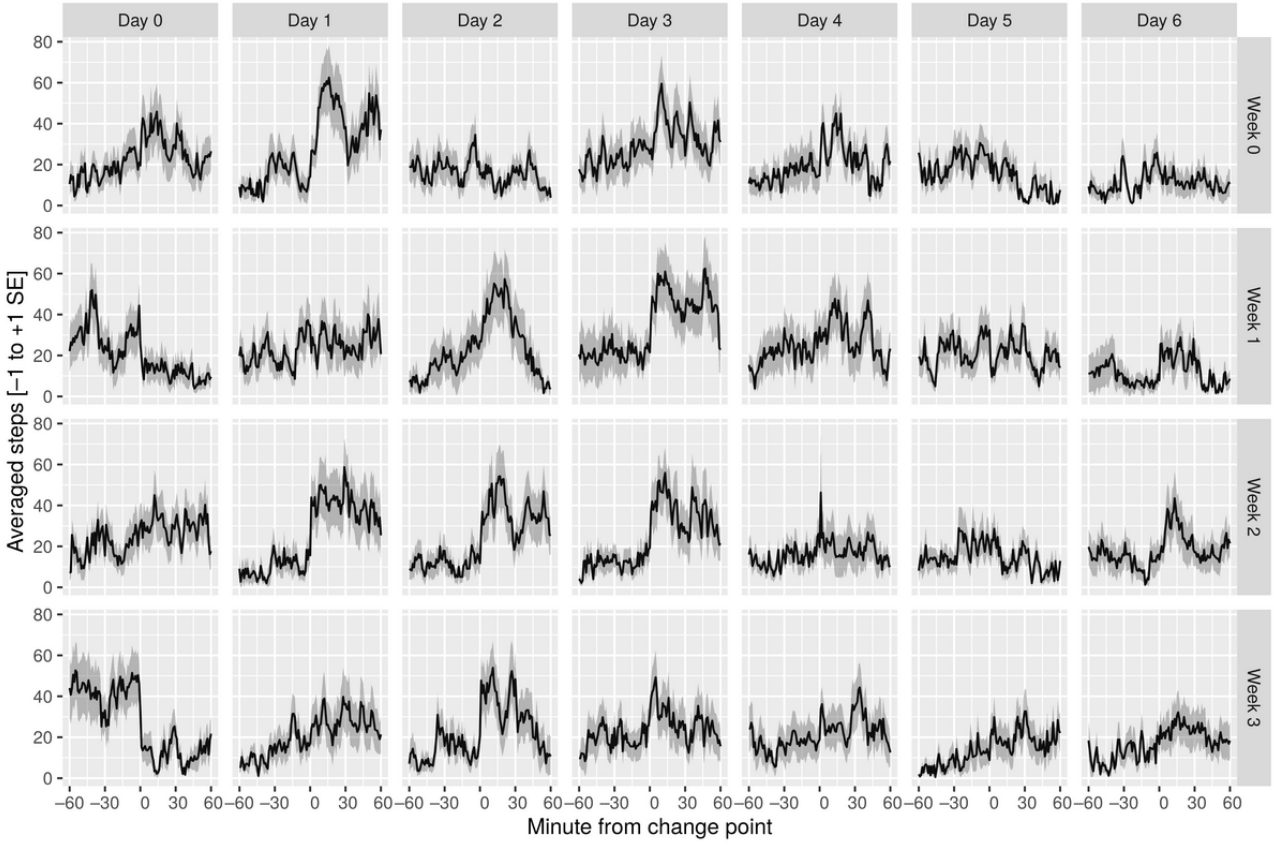
HR at 30 Minutes

The intervention resulted in a nonsignificant increase in HR compared with the baseline in the first week (t_{16}=1.15; *P*=.13; bootstrapped *P*=.13) and a significant increase in the second week (t_{16}=1.95; *P*=.03; bootstrapped *P*=.03). However, this effect diminished during week 3 (t_{16}=0.07; *P*=.47; bootstrapped *P*=.46), where we observed a nonsignificant decline compared with the baseline (Figure 6).

Step Count at 60 Minutes

Change-point analysis of step count revealed no clear linear trend (Table 3)—although we observed an insignificant decrease in the first week (t_{17}=−0.23; *P*=.59; bootstrapped *P*=.58), there was a significant increase in the second week (t_{17}=1.35; *P*=.09; bootstrapped *P*=.05). In week 3 (t_{17}=−0.21; *P*=.58; bootstrapped *P*=.56), we observed a nonsignificant decline compared with the baseline (see Figure 7 for the complete time course).

Figure 7. Change-point analysis: steps (mean+SE of the mean).



Step Count at 30 Minutes

As shown in Table 3, the change-point analysis of step count revealed a trend similar to the 60-minute measurement period—although we observed an insignificant increase in the first week (t_{17}=0.50; *P*=.30; bootstrapped *P*=.29), there was an increase in the second week, which was significant (t_{17}=1.34; *P*=.09; bootstrapped *P*=.05). In week 3 (t_{17}=−0.22; *P*=.58; bootstrapped *P*=.58), we observed a nonsignificant decline compared with the baseline (Figure 7).

Discussion

Principal Findings

This study, to our knowledge, is the first to develop a behavior change, theory-informed PA mHealth intervention with personalized prompts for adolescents aged 16 to 18 years evaluated using a within-subject experimental design. In contrast to the widespread 1D approach (eg, step count only [12,24,38,40]), this study involved the inclusion of 4 outcome measures to assess the multidimensional PA user profiles.

Overall, the results showed that the Mobile App for Physical Activity smartphone-based intervention produced significant reductions in sedentary time among adolescents during the first week of the trial. This trend, although it remained positive, diminished over time. This may be related to several reasons, including the holiday period, or certain aspects of the intervention being perceived as burdensome. This suggests that the implementation of the Mobile App for Physical Activity intervention may result in better health outcomes for adolescents, although there is currently insufficient evidence available to determine a specific dose-response relationship between sedentary time and health outcomes in adolescents [81]. Our findings indicate that the intervention had no effect on MET-based MVPA minutes, although the descriptive increase may give reason for further investigation. Although the results suggested no overall change in HR-based MVPA minutes, the results from the change-point analyses suggest that the personalized PA prompts significantly increased HR per minute

(bpm) during the second week of the study. There were no significant increases in the participants' overall step count; however, the personalized PA prompts resulted in a marginally significant increase in step counts per minute in the second week of the study. The results also revealed that the participants' engagement, based on the amount of missing data and responses to app suggestions, although initially high, decreased over the study period. These results may suggest that the intervention was successful in giving adolescents a nudge strong enough to interrupt and decrease their sedentary behavior but insufficient for a more high-effort increase in MVPA minutes. Personalized PA prompts, although moderately successful in promoting a light activity increase, did not result in a more intense MVPA minute increase.

As noted in the *Results* section, there were noticeable differences between participants from the 2 recruitment sites. This study was carried out between the summer term and the vacation period; therefore, this divergence may be explained by the fact that the 2 schools differed in their holiday calendars. Specifically, although the holiday period for students from Luxembourgish schools started at week 3 of the study, students from the international school were in the holiday period earlier, from week 1 onward. This may have resulted in an earlier decrease in PA for students from the international school compared with students from Luxembourgish schools. The descriptive lack of correlation between step count levels and MVPA levels confirmed the importance of accounting for various outcome measures while tracking the participants' PA in several dimensions.

Our study supports the results of a previous study by Bond et al [25], where a smartphone-based intervention yielded significant decreases in MET-based sedentary behavior in adults, which may confirm that smartphone-based PA interventions also have a high potential among adolescent populations. These results are also in line with some of the previous findings of the studies by Rabbi et al [23] and Klasnja et al [24] supporting the beneficial impact of personalized PA suggestions for adolescents. These results partially confirm the findings of the study by Kramer et al [82] supporting the use of financial incentives to initiate increased PA. However, future interventions should consider the exit strategy where, in time, participants would sustain increased PA levels based on intrinsic rather than extrinsic motivation. Challenges concerning the limited engagement of adolescents (based on the amount of missing data and decreased response to app suggestions over time) were similar to problems encountered by Lubans et al [83]. Engagement with mHealth PA interventions remains an important challenge to overcome for behavior change experts and developers in future interventions. Finally, our findings partially confirm the findings of the study by Gaudet et al [27], in which a minimalistic intervention based on the Fitbit activity tracker resulted in MVPA minute increases in adolescents, which may suggest that interventions including commercial fitness trackers may be advantageous for interventions among adolescent populations. Most published smartphone-based intervention studies such as ours include a relatively small participant sample. Therefore, it is important for future studies to replicate these findings and extend them to larger samples to further investigate approaches to increase adolescents' PA.

This study used a set of devices and a data platform that are designed to improve sedentary behavior and PA levels among adolescents and are currently commercially available. With their accessibility and relatively low price, compact and waterproof HR- and GPS-powered wrist-worn devices [31] in combination with research-grade data collection platforms provide researchers with attractive solutions for data collection and analysis, mitigating burdensomeness and intervention development time span.

A small number of studies on adolescents in the domain of mHealth PA and even fewer studies using theory-informed interventions call for future research in this area to further knowledge accumulation, both qualitative and quantitative. Systematic methods of intervention development with the help of tools such as the Behaviour Change Wheel and the BCT Taxonomy should be applied further by researchers to allow for the identification of effective intervention components and BCTs for the adolescent age group. Further sustainability of PA and sedentary behavior changes should be investigated via longitudinal studies. Finally, future research should implement alternative designs such as a within-subject design or MRT, which may investigate the efficacy of the intervention's individual components within a relatively short study duration.

These results suggest the feasibility and promise of smartphone-based PA interventions with personalized PA suggestions for adolescents. Although minimalist in nature, the introduction of such an intervention may represent a sufficient trigger for adolescents to decrease their sedentary behavior and increase their PA levels.

Strengths and Limitations

This study has several strengths. It is one of the few studies to develop and test an mHealth PA intervention for adolescents. Key methodological strengths include (1) the multidimensional PA profile assessment, specifically using versatile outcome measures; (2) the rigorous multistage theoretical development of the intervention guided by intervention development frameworks, taxonomies, and the latest research findings; and (3) the use of the latest wearable device and data collection platform, which presented inherent advantages and features, including undemanding data collection, quick device acceptance by participants, and prompt feedback time between participants and researchers.

This study also has important limitations. First, probably aggravated by restrictions imposed by the COVID-19 pandemic, we managed to enroll only a relatively small participant sample. Second, some participants did not wear the device for the entire duration of the study, taking it off during sleep or certain activities as wearing a watch or a fitness tracker was considered dangerous; for instance, in martial arts classes. The participants also forgot to wear the device on several occasions after sleep or forgot to charge the device in a timely manner, which resulted in missing data. Another important limitation is the short-term and small-scale nature of this study, which reduces the possibility to come to exhaustive conclusions. It is also

important to note that the participants took part in the study during the summer holiday period, which may have affected their PA patterns. In line with previous studies, during the baseline week of the study, we turned all the notifications off. However, we could not fully ensure that the participants would not use a Fitbit app or check the data provided by the fitness tracker either; similarly, we could not ascertain that the notifications during the treatment weeks would be read at once. The participants' second-language proficiency may have affected their overall engagement with the intervention. Finally, as the proprietary algorithms used to calculate HR- and MET-based MVPA minutes are not publicly available, caution must be taken when interpreting PA data collected by such trackers.

Despite these limitations, this study provides preliminary evidence of the usefulness of an mHealth PA smartphone intervention while shedding light on potential directions for future mHealth PA smartphone intervention developments.

Conclusions

This study provides preliminary evidence of the benefits of the Mobile App for Physical Activity intervention for modest yet significant reductions in the participants' sedentary time and the beneficial role of personalized PA prompts. These results also provide further evidence of the benefits and relative efficacy of personalized activity suggestions for inclusion in smartphone-based PA interventions. This study also provides an example of how to guide the development of subsequent smartphone-based mHealth PA interventions for adolescents. Future investigations should focus on replicating these findings and testing the potential for scalability of such an intervention in larger population samples.

Acknowledgments

AD was supported by an Industrial Fellowship from the Luxembourg National Research Fund (reference 12674722) and by Actimage Luxembourg SA. The authors would like to gratefully acknowledge their colleague and one of the coauthors of this paper, ASU, who helped immensely with the statistical analysis.

Conflicts of Interest

None declared.

References

1. Health at a Glance: Europe 2016 – State of Health in the EU Cycle. Organisation for Economic Co-operation and Development, European Union. 2016 Nov 23. URL: https://www.oecd-ilibrary.org/social-issues-migration-health/health-at-a-glance-europe-2016_9789264265592-en [accessed 2022-01-01]
2. Lee IM, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT, Lancet Physical Activity Series Working Group. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. Lancet 2012 Jul 21;380(9838):219-229 [FREE Full text] [doi: 10.1016/S0140-6736(12)61031-9] [Medline: 22818936]
3. Health at a Glance: Europe 2020 – State of Health in the EU Cycle. Organisation for Economic Co-operation and Development, European Union. 2020 Nov 19. URL: https://www.oecd-ilibrary.org/social-issues-migration-health/health-at-a-glance-europe-2020_82129230-en [accessed 2022-01-01]
4. A National Survey of Children and Young People's Physical Activity and Dietary Behaviours in New Zealand: 2008/09. National Institute for Health Innovation. 2010 Jun. URL: https://www.health.govt.nz/system/files/documents/publications/cyp-physical-activity-dietary-behaviours-08-09-keyfindgs.pdf [accessed 2022-01-01]
5. Patrick K, Griswold WG, Raab F, Intille SS. Health and the mobile phone. Am J Prev Med 2008 Aug;35(2):177-181 [FREE Full text] [doi: 10.1016/j.amepre.2008.05.001] [Medline: 18550322]
6. Klasnja P, Hekler EB, Shiffman S, Boruvka A, Almirall D, Tewari A, et al. Microrandomized trials: an experimental design for developing just-in-time adaptive interventions. Health Psychol 2015 Dec;34S:1220-1228 [FREE Full text] [doi: 10.1037/hea0000305] [Medline: 26651463]
7. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-Time Adaptive Interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
8. Hekler EB, Buman MP, Otten J, Castro CM, Grieco L, Marcus B, et al. Determining who responds better to a computer- vs. human-delivered physical activity intervention: results from the community health advice by telephone (CHAT) trial. Int J Behav Nutr Phys Act 2013 Sep 22;10:109 [FREE Full text] [doi: 10.1186/1479-5868-10-109] [Medline: 24053756]
9. King AC, Hekler EB, Castro CM, Buman MP, Marcus BH, Friedman RH, et al. Exercise advice by humans versus computers: maintenance effects at 18 months. Health Psychol 2014 Feb;33(2):192-196 [FREE Full text] [doi: 10.1037/a0030646] [Medline: 23421896]
10. King AC, Hekler EB, Grieco LA, Winter SJ, Sheats JL, Buman MP, et al. Harnessing different motivational frames via mobile phones to promote daily physical activity and reduce sedentary behavior in aging adults. PLoS One 2013 Apr 25;8(4):e62613 [FREE Full text] [doi: 10.1371/journal.pone.0062613] [Medline: 23638127]

11. Schoeppe S, Alley S, Van Lippevelde W, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 07;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
12. Walsh JC, Corbett T, Hogan M, Duggan J, McNamara A. An mHealth intervention using a smartphone app to increase walking behavior in young adults: a pilot study. JMIR Mhealth Uhealth 2016 Sep 22;4(3):e109 [FREE Full text] [doi: 10.2196/mhealth.5227] [Medline: 27658677]
13. Wang L, Miller LC. Just-in-the-Moment Adaptive Interventions (JITAI): a meta-analytical review. Health Commun 2020 Nov;35(12):1531-1544. [doi: 10.1080/10410236.2019.1652388] [Medline: 31488002]
14. Hardeman W, Houghton J, Lane K, Jones A, Naughton F. A systematic review of just-in-time adaptive interventions (JITAIs) to promote physical activity. Int J Behav Nutr Phys Act 2019 Apr 03;16(1):31 [FREE Full text] [doi: 10.1186/s12966-019-0792-7] [Medline: 30943983]
15. Knight E, Stuckey MI, Prapavessis H, Petrella RJ. Public health guidelines for physical activity: is there an app for that? A review of android and apple app stores. JMIR Mhealth Uhealth 2015 May 21;3(2):e43 [FREE Full text] [doi: 10.2196/mhealth.4003] [Medline: 25998158]
16. Stuckey MI, Carter SW, Knight E. The role of smartphones in encouraging physical activity in adults. Int J Gen Med 2017 Sep 12;10:293-303 [FREE Full text] [doi: 10.2147/IJGM.S134095] [Medline: 28979157]
17. Michie S, Carey RN, Johnston M, Rothman AJ, de Bruin M, Kelly MP, et al. From theory-inspired to theory-based interventions: a protocol for developing and testing a methodology for linking behaviour change techniques to theoretical mechanisms of action. Ann Behav Med 2018 May 18;52(6):501-512 [FREE Full text] [doi: 10.1007/s12160-016-9816-6] [Medline: 27401001]
18. Domin A, Spruijt-Metz D, Theisen D, Ouzzahra Y, Vögele C. Smartphone-based interventions for physical activity promotion: scoping review of the evidence over the last 10 years. JMIR Mhealth Uhealth 2021 Jul 21;9(7):e24308 [FREE Full text] [doi: 10.2196/24308] [Medline: 34287209]
19. McCallum C, Rooksby J, Gray CM. Evaluating the impact of physical activity apps and wearables: interdisciplinary review. JMIR Mhealth Uhealth 2018 Mar 23;6(3):e58 [FREE Full text] [doi: 10.2196/mhealth.9054] [Medline: 29572200]
20. Pham Q, Wiljer D, Cafazzo JA. Beyond the randomized controlled trial: a review of alternatives in mHealth clinical trial methods. JMIR Mhealth Uhealth 2016 Sep 09;4(3):e107 [FREE Full text] [doi: 10.2196/mhealth.5720] [Medline: 27613084]
21. Kumar S, Nilsen WJ, Abernethy A, Atienza A, Patrick K, Pavel M, et al. Mobile health technology evaluation: the mHealth evidence workshop. Am J Prev Med 2013 Aug;45(2):228-236 [FREE Full text] [doi: 10.1016/j.amepre.2013.03.017] [Medline: 23867031]
22. Riley WT, Glasgow RE, Etheredge L, Abernethy AP. Rapid, responsive, relevant (R3) research: a call for a rapid learning health research enterprise. Clin Transl Med 2013 May 10;2(1):10 [FREE Full text] [doi: 10.1186/2001-1326-2-10] [Medline: 23663660]
23. Rabbi M, Pfammatter A, Zhang M, Spring B, Choudhury T. Automated personalized feedback for physical activity and dietary behavior change with mobile phones: a randomized controlled trial on adults. JMIR Mhealth Uhealth 2015 May 14;3(2):e42 [FREE Full text] [doi: 10.2196/mhealth.4160] [Medline: 25977197]
24. Klasnja P, Smith S, Seewald NJ, Lee A, Hall K, Luers B, et al. Efficacy of contextually tailored suggestions for physical activity: a micro-randomized optimization trial of HeartSteps. Ann Behav Med 2019 May 03;53(6):573-582 [FREE Full text] [doi: 10.1093/abm/kay067] [Medline: 30192907]
25. Bond DS, Thomas JG, Raynor HA, Moon J, Sieling J, Trautvetter J, et al. B-MOBILE--a smartphone-based intervention to reduce sedentary time in overweight/obese individuals: a within-subjects experimental trial. PLoS One 2014 Jun 25;9(6):e100821 [FREE Full text] [doi: 10.1371/journal.pone.0100821] [Medline: 24964010]
26. Kramer JN, Künzler F, Mishra V, Presset B, Kotz D, Smith S, et al. Investigating intervention components and exploring states of receptivity for a smartphone app to promote physical activity: protocol of a microrandomized trial. JMIR Res Protoc 2019 Jan 31;8(1):e11540 [FREE Full text] [doi: 10.2196/11540] [Medline: 30702430]
27. Gaudet J, Gallant F, Bélanger M. A bit of fit: minimalist intervention in adolescents based on a physical activity tracker. JMIR Mhealth Uhealth 2017 Jul 06;5(7):e92 [FREE Full text] [doi: 10.2196/mhealth.7647] [Medline: 28684384]
28. Michie S, Atkins L, West R. The Behaviour Change Wheel: A Guide To Designing Interventions. Sutton, UK: Silverback Publishing; 2014.
29. Tudor-Locke C, Craig CL, Beets MW, Belton S, Cardon GM, Duncan S, et al. How many steps/day are enough? For children and adolescents. Int J Behav Nutr Phys Act 2011 Jul 28;8:78 [FREE Full text] [doi: 10.1186/1479-5868-8-78] [Medline: 21798014]
30. Fitbit application. Google Play. 2021. URL: https://play.google.com/store/apps/details?id=com.fitbit.FitbitMobile&hl=en_IN&gl=US [accessed 2022-01-01]
31. Fitbit Charge 4. Fitbit Inc. 2021. URL: https://www.fitbit.com/global/us/products/trackers/charge4 [accessed 2022-01-01]
32. Thompson D, Peacock O, Western M, Batterham AM. Multidimensional physical activity: an opportunity, not a problem. Exerc Sport Sci Rev 2015 Apr;43(2):67-74 [FREE Full text] [doi: 10.1249/JES.0000000000000039] [Medline: 25607280]
33. Actigraph device comparison. ActiGraph. 2020. URL: https://actigraphcorp.com/activity-monitor-comparison/ [accessed 2022-01-01]

34. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 2015 Dec 18;12:159 [FREE Full text] [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]
35. Straiton N, Alharbi M, Bauman A, Neubeck L, Gullick J, Bhindi R, et al. The validity and reliability of consumer-grade activity trackers in older, community-dwelling adults: a systematic review. Maturitas 2018 Jun;112:85-93. [doi: 10.1016/j.maturitas.2018.03.016] [Medline: 29704922]
36. Choi J, Lee JH, Vittinghoff E, Fukuoka Y. mHealth physical activity intervention: a randomized pilot study in physically inactive pregnant women. Matern Child Health J 2016 May;20(5):1091-1101 [FREE Full text] [doi: 10.1007/s10995-015-1895-7] [Medline: 26649879]
37. Duncan MJ, Vandelanotte C, Trost SG, Rebar AL, Rogers N, Burton NW, et al. Balanced: a randomised trial examining the efficacy of two self-monitoring methods for an app-based multi-behaviour intervention to improve physical activity, sitting and sleep in adults. BMC Public Health 2016 Jul 30;16:670 [FREE Full text] [doi: 10.1186/s12889-016-3256-x] [Medline: 27473327]
38. Korinek EV, Phatak SS, Martin CA, Freigoun MT, Rivera DE, Adams MA, et al. Adaptive step goals and rewards: a longitudinal growth model of daily steps for a smartphone-based walking intervention. J Behav Med 2018 Feb;41(1):74-86. [doi: 10.1007/s10865-017-9878-3] [Medline: 28918547]
39. Middelweerd A, Te Velde SJ, Mollee JS, Klein MC, Brug J. App-based intervention combining evidence-based behavior change techniques with a model-based reasoning system to promote physical activity among young adults (Active2Gether): descriptive study of the development and content. JMIR Res Protoc 2018 Dec 21;7(12):e185 [FREE Full text] [doi: 10.2196/resprot.7169] [Medline: 30578198]
40. van Woudenberg TJ, Bevelander KE, Burk WJ, Smit CR, Buijs L, Buijzen M. A randomized controlled trial testing a social network intervention to promote physical activity among adolescents. BMC Public Health 2018 Apr 23;18(1):542 [FREE Full text] [doi: 10.1186/s12889-018-5451-4] [Medline: 29685112]
41. Fuller D, Colwell E, Low J, Orychock K, Tobin MA, Simango B, et al. Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: systematic review. JMIR Mhealth Uhealth 2020 Sep 08;8(9):e18694 [FREE Full text] [doi: 10.2196/18694] [Medline: 32897239]
42. Brewer W, Swanson BT, Ortiz A. Validity of Fitbit's active minutes as compared with a research-grade accelerometer and self-reported measures. BMJ Open Sport Exerc Med 2017 Sep 13;3(1):e000254 [FREE Full text] [doi: 10.1136/bmjsem-2017-000254] [Medline: 29018543]
43. Dorn D, Gorzelitz J, Gangnon R, Bell D, Koltyn K, Cadmus-Bertram L. Automatic identification of physical activity type and duration by wearable activity trackers: a validation study. JMIR Mhealth Uhealth 2019 May 23;7(5):e13547 [FREE Full text] [doi: 10.2196/13547] [Medline: 31124470]
44. Nazari G, MacDermid JC, Sinden KE, Richardson J, Tang A. Inter-instrument reliability and agreement of Fitbit charge measurements of heart rate and activity at rest, during the modified Canadian aerobic fitness test, and in recovery. Physiother Can 2019;71(3):197-206 [FREE Full text] [doi: 10.3138/ptc.2018-25] [Medline: 31719715]
45. Degroote L, De Paepe A, De Bourdeaudhuij I, Van Dyck D, Crombez G. Effectiveness of the mHealth intervention 'MyDayPlan' to increase physical activity: an aggregated single case approach. Int J Behav Nutr Phys Act 2021 Jul 07;18(1):92 [FREE Full text] [doi: 10.1186/s12966-021-01163-2] [Medline: 34233718]
46. King AC, Campero I, Sheats JL, Castro Sweet CM, Garcia D, Chazaro A, et al. Testing the comparative effects of physical activity advice by humans vs. computers in underserved populations: the COMPASS trial design, methods, and baseline characteristics. Contemp Clin Trials 2017 Oct;61:115-125 [FREE Full text] [doi: 10.1016/j.cct.2017.07.020] [Medline: 28739541]
47. Sullivan AN, Lachman ME. Behavior change with fitness technology in sedentary adults: a review of the evidence for increasing physical activity. Front Public Health 2017 Jan 11;4:289 [FREE Full text] [doi: 10.3389/fpubh.2016.00289] [Medline: 28123997]
48. Sniehotta FF, Schwarzer R, Scholz U, Schüz B. Action planning and coping planning for long-term lifestyle change: theory and assessment. Eur J Soc Psychol 2005 Jul;35(4):565-576. [doi: 10.1002/ejsp.258]
49. Bidargaddi N, Almirall D, Murphy S, Nahum-Shani I, Kovalcik M, Pituch T, et al. To prompt or not to prompt? A microrandomized trial of time-varying push notifications to increase proximal engagement with a mobile health app. JMIR Mhealth Uhealth 2018 Nov 29;6(11):e10123 [FREE Full text] [doi: 10.2196/10123] [Medline: 30497999]
50. WHO guidelines on physical activity and sedentary behaviour. World Health Organization. 2020 Nov 25. URL: https://www.who.int/publications/i/item/9789240015128 [accessed 2022-01-01]
51. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]
52. Mann T, de Ridder D, Fujita K. Self-regulation of health behavior: social psychological approaches to goal setting and goal striving. Health Psychol 2013 May;32(5):487-498. [doi: 10.1037/a0028533] [Medline: 23646832]
53. Walton A, Nahum-Shani I, Crosby L, Klasnja P, Murphy S. Optimizing digital integrated care via micro-randomized trials. Clin Pharmacol Ther 2018 Jul;104(1):53-58 [FREE Full text] [doi: 10.1002/cpt.1079] [Medline: 29604043]

54. Rabin C, Bock B. Desired features of smartphone applications promoting physical activity. Telemed J E Health 2011 Dec;17(10):801-803. [doi: 10.1089/tmj.2011.0055] [Medline: 22010977]
55. Chang TR, Kaasinen E, Kaipainen K. What influences users' decisions to take apps into use?: a framework for evaluating persuasive and engaging design in mobile Apps for well-being. In: Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia. 2012 Presented at: MUM '12; December 4-6, 2012; Ulm, Germany p. 1-10. [doi: 10.1145/2406367.2406370]
56. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013 Apr 18;15(4):e86 [FREE Full text] [doi: 10.2196/jmir.2583] [Medline: 23598614]
57. Ehlers DK, Huberty JL. Middle-aged women's preferred theory-based features in mobile physical activity applications. J Phys Act Health 2014 Sep;11(7):1379-1385. [doi: 10.1123/jpah.2012-0435] [Medline: 24368818]
58. Gowin M, Cheney M, Gwin S, Franklin Wann T. Health and fitness app use in college students: a qualitative study. Am J Health Educ 2015 Jul 06;46(4):223-230. [doi: 10.1080/19325037.2015.1044140]
59. Middelweerd A, van der Laan DM, van Stralen MM, Mollee JS, Stuij M, te Velde SJ, et al. What features do Dutch university students prefer in a smartphone application for promotion of physical activity? A qualitative approach. Int J Behav Nutr Phys Act 2015 Mar 01;12:31 [FREE Full text] [doi: 10.1186/s12966-015-0189-1] [Medline: 25889577]
60. Miyamoto SW, Henderson S, Young HM, Pande A, Han JJ. Tracking health data is not enough: a qualitative exploration of the role of healthcare partnerships and mHealth technology to promote physical activity and to sustain behavior change. JMIR Mhealth Uhealth 2016 Jan 20;4(1):e5 [FREE Full text] [doi: 10.2196/mhealth.4814] [Medline: 26792225]
61. Baretta D, Perski O, Steca P. Exploring users' experiences of the uptake and adoption of physical activity apps: longitudinal qualitative study. JMIR Mhealth Uhealth 2019 Feb 08;7(2):e11636 [FREE Full text] [doi: 10.2196/11636] [Medline: 30735143]
62. Lyons EJ, Lewis ZH, Mayrsohn BG, Rowland JL. Behavior change techniques implemented in electronic lifestyle activity monitors: a systematic content analysis. J Med Internet Res 2014 Aug 15;16(8):e192 [FREE Full text] [doi: 10.2196/jmir.3469] [Medline: 25131661]
63. Brannon EE, Cushing CC. A systematic review: is there an app for that? Translational science of pediatric behavior change for physical activity and dietary interventions. J Pediatr Psychol 2015 May;40(4):373-384. [doi: 10.1093/jpepsy/jsu108] [Medline: 25502745]
64. Williams SL, French DP. What are the most effective intervention techniques for changing physical activity self-efficacy and physical activity behaviour--and are they the same? Health Educ Res 2011 Apr;26(2):308-322. [doi: 10.1093/her/cyr005] [Medline: 21321008]
65. Olander EK, Fletcher H, Williams S, Atkinson L, Turner A, French DP. What are the most effective techniques in changing obese individuals' physical activity self-efficacy and behaviour: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2013 Mar 03;10:29 [FREE Full text] [doi: 10.1186/1479-5868-10-29] [Medline: 23452345]
66. O'Brien N, McDonald S, Araújo-Soares V, Lara J, Errington L, Godfrey A, et al. The features of interventions associated with long-term effectiveness of physical activity interventions in adults aged 55-70 years: a systematic review and meta-analysis. Health Psychol Rev 2015;9(4):417-433. [doi: 10.1080/17437199.2015.1012177] [Medline: 25689096]
67. Samdal GB, Eide GE, Barth T, Williams G, Meland E. Effective behaviour change techniques for physical activity and healthy eating in overweight and obese adults; systematic review and meta-regression analyses. Int J Behav Nutr Phys Act 2017 Mar 28;14(1):42 [FREE Full text] [doi: 10.1186/s12966-017-0494-y] [Medline: 28351367]
68. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
69. Bort-Roig J, Gilson ND, Puig-Ribera A, Contreras RS, Trost SG. Measuring and influencing physical activity with smartphone technology: a systematic review. Sports Med 2014 May;44(5):671-686. [doi: 10.1007/s40279-014-0142-5] [Medline: 24497157]
70. Dute DJ, Bemelmans WJ, Breda J. Using mobile apps to promote a healthy lifestyle among adolescents and students: a review of the theoretical basis and lessons learned. JMIR Mhealth Uhealth 2016 May 05;4(2):e39 [FREE Full text] [doi: 10.2196/mhealth.3559] [Medline: 27150850]
71. Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. Health Psychol 2008 May;27(3):379-387. [doi: 10.1037/0278-6133.27.3.379] [Medline: 18624603]
72. Michie S, Ashford S, Sniehotta FF, Dombrowski SU, Bishop A, French DP. A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: the CALO-RE taxonomy. Psychol Health 2011 Nov;26(11):1479-1498. [doi: 10.1080/08870446.2010.540664] [Medline: 21678185]
73. Michie S, Wood CE, Johnston M, Abraham C, Francis JJ, Hardeman W. Behaviour change techniques: the development and evaluation of a taxonomic method for reporting and describing behaviour change interventions (a suite of five studies involving consensus methods, randomised controlled trials and analysis of qualitative data). Health Technol Assess 2015 Nov;19(99):1-188 [FREE Full text] [doi: 10.3310/hta19990] [Medline: 26616119]
74. Carver CS, Scheier MF. On the Self-Regulation of Behavior. Cambridge, UK: Cambridge University Press; 1998.

75. Morton K, Sutton S, Hardeman W, Troughton J, Yates T, Griffin S, et al. A text-messaging and pedometer program to promote physical activity in people at high risk of type 2 diabetes: the development of the PROPELS follow-on support program. JMIR Mhealth Uhealth 2015 Dec 15;3(4):e105 [FREE Full text] [doi: 10.2196/mhealth.5026] [Medline: 26678750]
76. Domin A, Ouzzahra Y, Vögele C. Features and components preferred by adolescents in smartphone apps for the promotion of physical activity: focus group study. JMIR Hum Factors 2022;9(2):e33972 [FREE Full text] [doi: 10.2196/33972]
77. Kowalski KC, Crocker PR, Donen RM. The Physical Activity Questionnaire for Older Children (PAQ-C) and Adolescents (PAQ-A) Manual. University of Saskatchewan. 2004 Aug. URL: https://www.prismsports.org/UserFiles/file/PAQ_manual_ScoringandPDF.pdf [accessed 2022-01-01]
78. Biddle SJ, Gorely T, Pearson N, Bull FC. An assessment of self-reported physical activity instruments in young people for population surveillance: project ALPHA. Int J Behav Nutr Phys Act 2011 Jan 02;8:1 [FREE Full text] [doi: 10.1186/1479-5868-8-1] [Medline: 21194492]
79. What is Fitabase? Small Steps Labs. 2021. URL: https://app.fitabase.com/ [accessed 2022-01-01]
80. About RStudio. RStudio. 2021. URL: https://www.rstudio.com/about/ [accessed 2022-01-01]
81. Chaput JP, Willumsen J, Bull F, Chou R, Ekelund U, Firth J, et al. 2020 WHO guidelines on physical activity and sedentary behaviour for children and adolescents aged 5-17 years: summary of the evidence. Int J Behav Nutr Phys Act 2020 Nov 26;17(1):141 [FREE Full text] [doi: 10.1186/s12966-020-01037-z] [Medline: 33239009]
82. Kramer JN, Künzler F, Mishra V, Smith SN, Kotz D, Scholz U, et al. Which components of a smartphone walking app help users to reach personalized step goals? Results from an optimization trial. Ann Behav Med 2020 Jun 12;54(7):518-528 [FREE Full text] [doi: 10.1093/abm/kaaa002] [Medline: 32182353]
83. Lubans DR, Smith JJ, Skinner G, Morgan PJ. Development and implementation of a smartphone application to promote physical activity and reduce screen-time in adolescent boys. Front Public Health 2014 May 20;2:42 [FREE Full text] [doi: 10.3389/fpubh.2014.00042] [Medline: 24904909]

Abbreviations

BCT: behavior change technique
HR: heart rate
JITAI: just-in-time adaptive intervention
MET: metabolic equivalent of task
mHealth: mobile health
MRT: microrandomized trial
MVPA: moderate to vigorous physical activity
PA: physical activity
PAQ-A: Physical Activity Questionnaire for Adolescents
RCT: randomized controlled trial
WHO: World Health Organization

Edited by A Mavragani; submitted 22.11.21; peer-reviewed by A Middelweerd, L Wang; comments to author 02.02.22; revised version received 13.03.22; accepted 11.04.22; published 10.06.22.

Please cite as:
Domin A, Uslu A, Schulz A, Ouzzahra Y, Vögele C
A Theory-Informed, Personalized mHealth Intervention for Adolescents (Mobile App for Physical Activity): Development and Pilot Study
JMIR Form Res 2022;6(6):e35118
URL: https://formative.jmir.org/2022/6/e35118
doi:10.2196/35118
PMID:35687409

Original Paper

Digital Health Solutions and State of Interoperability: Landscape Analysis of Sierra Leone

Emeka Chukwu[1], BEng, MSc, CGEIT; Lalit Garg[1], BE, MSc, PhD; Edward Foday[2], BA, MSc, MPhil; Abdul Konomanyi[3], BEng, MSc; Royston Wright[4], BSc, MA, MSc; Francis Smart[2], MD, MPH

[1]Department of Computer Information Systems, Faculty of Information and Communication Technology, University of Malta, Msida, Malta

[2]Directorate of Planning, Policy, and Information, Ministry of Health and Sanitation, Freetown, Sierra Leone

[3]Directorate of eGovernment, Ministry of Information and Communication, Freetown, Sierra Leone

[4]Monitoring and Evaluation Unit, Health and Nutrition, UNICEF, Freetown, Sierra Leone

Corresponding Author:
Emeka Chukwu, BEng, MSc, CGEIT
Department of Computer Information Systems
Faculty of Information and Communication Technology
University of Malta
PG room A24, Level 0
Msida, MSD2080
Malta
Phone: 356 99330888
Email: nnaemeka_ec@hotmail.com

Abstract

Background: The government and partners have invested heavily in the health information system (HIS) for service delivery, surveillance, reporting, and monitoring. Sierra Leone's government launched its first digital health strategy in 2018. In 2019, a broader national innovation and digital strategy was launched. The health pillar direction will use big data and artificial intelligence (AI) to improve health care in general and maternal and child health in particular. Understanding the number, distribution, and interoperability of digital health solutions is crucial for successful implementation strategies.

Objective: This paper presents the state of digital health solutions in Sierra Leone and how these solutions currently interoperate. This study further presents opportunities for big data and AI applications.

Methods: All the district health management teams, all digital health implementing organizations, and a stratified sample of 72 (out of 1284) health facilities were purposefully selected from all health districts and surveyed.

Results: The National Health Management Information System's (NHMIS's) aggregate reporting solution populated by health facility forms HF1 to HF9 was, by far, the most used tool. A health facility–based weekly aggregate electronic integrated disease surveillance and response solution was also widely used. Half of the health facilities had more than 2 digital health solutions in use. The different digital health software solutions do not share data among one another, though aggregate reporting data were sent as necessary. None of the respondents use any of the health care registries for patient, provider, health facility, or terminology identification.

Conclusions: Many digital health solutions are currently used at health facilities in Sierra Leone. The government can leverage current investment in HIS from surveillance and reporting for using big data and AI for care. The vision of using big data for health care is achievable if stakeholders prioritize individualized and longitudinal patient data exchange using agreed use cases from national strategies. This study has shown evidence of distribution, types, and scale of digital health solutions in health facilities and opportunities for leveraging big data to fill critical gaps necessary to achieve the national digital health vision.

(JMIR Form Res 2022;6(6):e29930) doi:10.2196/29930

KEYWORDS

digital health; mHealth; mobile health; eHealth; Health information and communication technologies; Sierra Leone; big data; HIE; interoperability

Introduction

Background

The Ministry of Health and Sanitation (MoHS) inaugurated an eHealth coordination hub in 2017 to facilitate the systematic application of digital health solutions (or services and applications) for health systems improvement through data [1]. This culminated in the launch of the first national digital health strategy for 2018-2023 [2]. In 2019, the Directorate of Science Technology and Innovation at the Presidency also launched a broader National Innovation and Digital Strategy for 2019-2029. The broader strategy set out three strategic health pillars [3]:

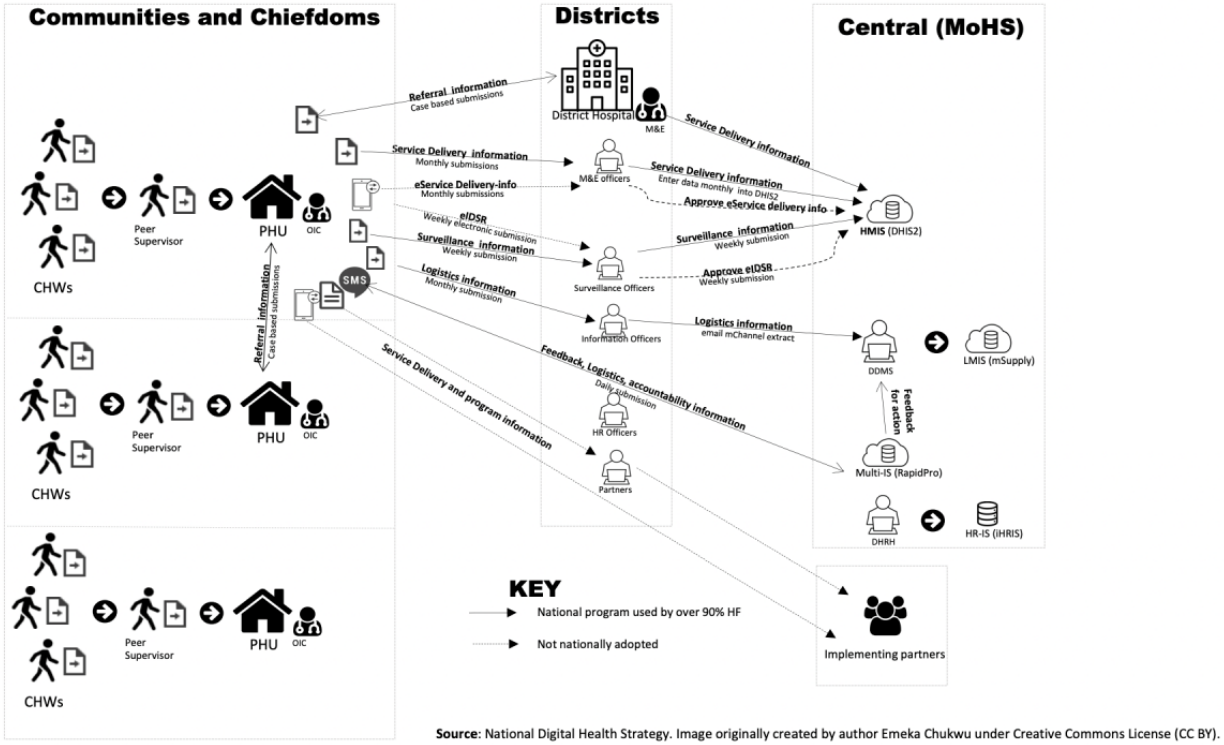
1. Application of data science methods (including artificial intelligence [AI]) to diagnostic images, genomics, mobility, environmental, and other data analytical methods for automated disease diagnostics, predicting disease outbreaks, disease prevention, and identifying high-risk groups for planning and resource allocation
2. Use of AI to support junior-level and expert-level health care practitioners to make better health care decisions related to treatments and referrals in quicker time and for more people
3. Use of an integrated community and technology approach to significantly reduce maternal and child mortality.

As seen from the strategic health pillars, big data and AI are fundamental to the success of the national visions. As noted by Andreu-Perez et al [4], the phrase "Big data" is becoming a buzzword whose usage continues to double every year. They went on to outline the six Vs of big data in health care: value, volume, velocity, variety, veracity, and variability. Health care data's clinical and public health *value* in financial terms and in health outcomes continue to drive interests in health care data globally. For instance, the total UK Health and Social Care expenditure in 2020 was 140 billion pounds representing 4% of total expenditure [5]. Similarly, the per capita national health expenditure in the United States stood at US $10,739 as of 2017 [6]. Low- and middle-income countries such as Sierra Leone continue to underperform their high-income counterparts in terms of health outcomes. This is particularly evident in the maternal and child health outcomes [7].

The *volume* of health care data generated in the United States alone in 2011 was 150 exabytes (10^{18} gigabytes) [8]. While we do not currently have such statistics for Sierra Leone, this helps to indicate the volume of data expected in a truly big data–enabled health system. One can expect individualized health care data such as vital signs, historical information, or high-resolution sound and imaging data to be the bedrock of a big data–enabled health system. The different types of data coming from different sources such as sensors, mobile apps, and hospital clinic information systems all represent the *variety* component of a big data system. Genetic, laboratory, and population health data all introduce their different challenges and perspectives to health care data. There is now increasing demand for real-time health care data: velocity. Health systems strive to improve data quality as the data travel from source to where they are used: Veracity. The quality of data is dependent on data sources, and the quality reduces with each human interface. Will a given datum be available over time? What is the guarantee of its consistency? These and other questions are addressed in the variability and trustworthiness of health care data. In Sierra Leone, the government has improved health care data reporting completeness and timeliness of reporting from the health facilities in the country.

There are 1284 health facilities in Sierra Leone, including 24 district hospitals, and the rest are primary health care units (PHUs) [9]. Each health facility service delivery point and disease surveillance unit reports through the District Health Information System (DHIS) to the MoHS at the central level [9,10]. Data flow through the different levels of information is as shown in Figure 1 from the national digital health strategy [2]. The health facility's aggregate service delivery report submission rate was 98.6%, with 91.4% submitting on time [2].

Figure 1. Health Information flow architecture. CHC: community health Center; CHW: community health worker; DDMS: DDMS - Directorate of Drugs and Medical Supplies; DHIS2: District Health Information Software; DHRH: District Human Resource for Health; eIDSR: electronic Integrated Disease Surveillance and Response; HF: Health Facility; HMIS: Health Management Information System; HR: human resources; HR-IS (iHRIS): Human Resources Information System; LMIS: Logistics Management Information System; M&E: Monitoring and Evaluation; MCHP: maternal and child health posts; MoHS: Ministry of Health and Sanitation; OIC: Officer in Charge; PHU: Primary Healthcare Unit.



Source: National Digital Health Strategy. Image originally created by author Emeka Chukwu under Creative Commons License (CC BY).

Study Objective

This study was commissioned to understand the different health facilities and other health systems' digital health solutions. We also analyzed and present how these solutions are used to share information among different stakeholders in the health system, the structure of data shared, and how the data are used for decision-making. Findings from this mapping exercise conducted in 2019 provide evidence of the linkage between the availability of individualized (or longitudinal) digital health data and the exchange of these data in support of the national digital health vision. The methodology section discusses the overall investigation methodologies, including sampling, data collection, analysis, and interpretation. Next, we present our findings and the implications with key recommendations.

Methods

Survey Tools

For this survey, the 13 district health medical officers (DMOs) were targeted for survey. The DMOs are the health care policy implementers in their respective district and oversee district health programs at health facilities in their district. In addition, a stratified sample of 72 health facilities were also selected and visited. A separate tool was developed for digital health implementing organizations in the country. The 3 survey tools used for the survey are those shown in Table 1. Each of the questionnaires were coded into CommCare electronic mobile form [11]. Each DMO was visited, and the questionnaire was applied. Similarly, the digital health implementing organizations were visited, and a questionnaire was applied.

Table 1. Survey tools used for digital health mapping.

Tool name	Target respondent	Alternative	Where it is applied
Assessment Survey for District Health Management Team	District health medical officer	Authorized representative	District level
Assessment Survey for Implementing Partners	Implementing partners or implementing ministries, departments and agencies leads	Authorized representative	National or district level
Health Facility Checklist	Hospital superintendent or PHU[a] officer in charge	Representative	Health facility

[a]PHU: primary health care unit.

Health Facility Sampling

For the health facilities, a stratified sampling technique that includes 72 health facilities (out of 1284) with a confidence level of 95% and 11% margin of error to ensure findings can be generalizable. Health facilities were stratified into urban and rural and into high, medium, or low digital health activity health facilities, using information from the Directorate of Policy, Planning and Information (DPPI) at the MoHS working with respective DMOs. The health facilities surveyed include 17 urban and 55 rural health facilities, as shown in Table 2. In total, 96% (n=69) are public sector health facilities. The district's DMO determines urban-rural classification. A facility is classified as low digital health activity if no digital health solution is in use at the health facility, medium if 1 or 2 solutions are used, and high if 3 or more.

These classifications were in addition to their Hospital versus PHU categorizations. In order to arrive at our sample size, a minimum of 5 health facilities were purposefully targeted for selection in each district visited. Each district DMO suggested one district hospital as part of the 5 survey health facilities. One health facility with high digital health activity was prioritized, followed by one with medium activity, followed by low (or no) activity. The process outlined above is repeated until the required number of health facilities is reached.

Table 2. Distribution of health facilities surveyed, by district (health facility survey).

Health facilities by district, n																							
Bo		Bombali		Bonthe		Kailahun		Kambia	Kenema		Koinadugu	Kono		Moyamba		Port loko		Pujehun		Tonkolili		Western Rural	Western Urban
R[a]	U[b]	R	U	R	U	R	U	R	U	R	U	R	U	R	U	R	U	R	U	R	U	R	R
4	2	4	1	4	2	4	1	5	3	1	5	4	1	5	1	5	4	1	4	1	4		6

[a]R: rural.

[b]U: urban.

DMO and Implementer Sampling

In addition to health facilities, all the DMOs and all identified digital health implementers were surveyed. Implementing partner organizations were included for the structured survey if they have an active digital health implementation at the national or district level, as determined by the DPPI at the MoHS. The implementer survey tool covered the state of their digital health solutions. In total, 15 implementing organizations reported supporting digital health solutions in Sierra Leone and were all surveyed. Each implementer had one or more digital health solutions at various degrees of implementation. Similarly, the DMO—heading the District Health Management Team (DHMT)—was surveyed for the state of digital health solutions at the health facility they oversee.

Data Collection and Analysis

Study personnel surveyed targeted respondents at the national level and then moved to the district and health facility levels. No identifiable information was collected as authorized institutional representatives were surveyed. The quantitative data collection and structured interviews were carried out using the CommCare mobile app, which facilitated automatic data transmission to the cloud for easy access. Enumerators collected data using mobile forms, which were aggregated into a Microsoft Excel spreadsheet. The aggregated data were later analyzed with "pandas" and "matplotlib" libraries of Python.

Results

Here we present our study findings concerning the state of digital health solutions in Sierra Leone. The digital health solutions group the findings, data sharing practices, and current data use.

Solutions (Services and Applications)

DHMT Survey Findings

The number and distribution from the survey is presented as reported by the DHMT and by health facilities. Based on the survey of DHMTs, Kailahun, Kenema, Karane, Pujehun, Moyamba, Freetown-Western-Rural, and Freetown-Western-Urban reported having 4 or more digital health services and applications. Bo and Kono districts had 3, and the remaining districts had 2 or fewer. Among the digital health solutions, every district used the national District Health Information Software (DHIS2).

Health Facility Survey Findings

Similarly, the health facility survey showed that in total, 3 health facilities had 4 or more digital health services and applications in use, 4 facilities had 3 solutions in use, and the majority had 2 services and applications in use or only the DHIS for aggregate reporting. Figure 2 provides a breakdown of this distribution by hospitals and the different PHUs. Based on health facility respondents, digital tools used by health facilities were aggregate electronic Integrated Disease Surveillance (eIDSR) and DHIS2. Others include the case-based reporting tool based on odk, commcare, ihris, vaxtrac, and healthConnect. Some health facilities also indicated using SMS reporting through RapidPro and the NHMIS-paper-form HF1_HF9 reporting tool.

Almost all facilities reported that the services and applications were functional, except a negligible few, which were reported not to be working at the time of the data collector's visit. The World Health Organization (WHO) classified digital health interventions into client, health care provider, health system administrator, and data services–facing services and applications (solutions) [12]. Based on these categorizations, the majority of the services and applications deployed were either for data services or for health care providers. Table 2 shows the

distribution and purpose of these digital health solutions by health facility type. SA_1 represents the first service and application, and SA_5 represents the fifth service and application. The first column in Table 3 implies that 5 hospitals had at least 1 service and application that are used for data service. In addition, there is one hospital whose fourth service and application are used for data services, and one hospital whose fifth service and application are used for data services.

Respondents at the health facilities surveyed were asked about how the services and applications were accessed at their facilities. The hospitals accessed their digital health solutions mainly using computers and through the internet. Similarly, the PHUs accessed their digital health solutions primarily using tablets (or smartphones) (Figure 3), although the maternal and child health posts used more basic phones than the other PHUs on average. Each rectangle represents access type, relative sizes of each rectangle indicate the number of health facilities, and the color indicates the health facility type.

Figure 2. Health facilities by the number of digital health activities (health facility survey). CHC: community health centerl; CHP: community health post; MCHP: maternal and child health post; No: number.

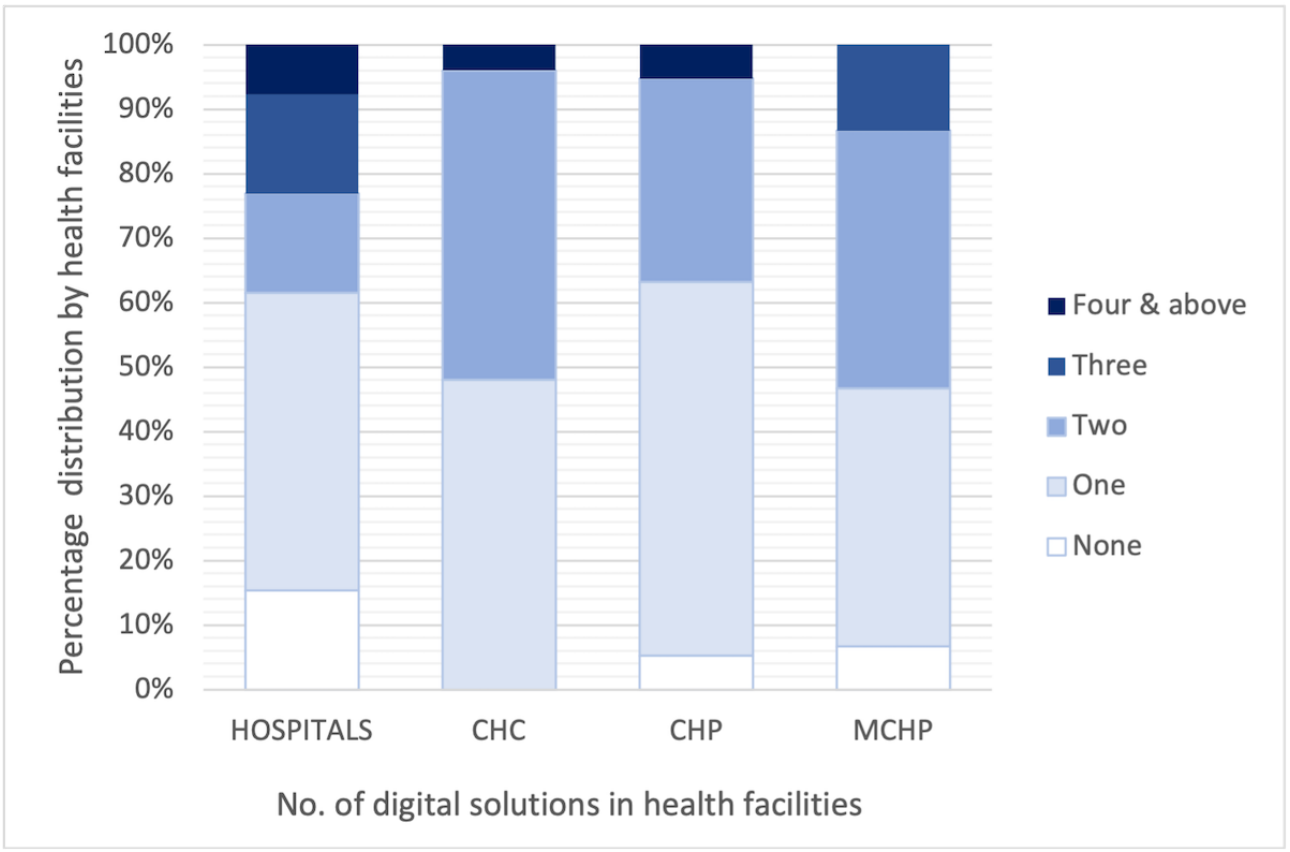
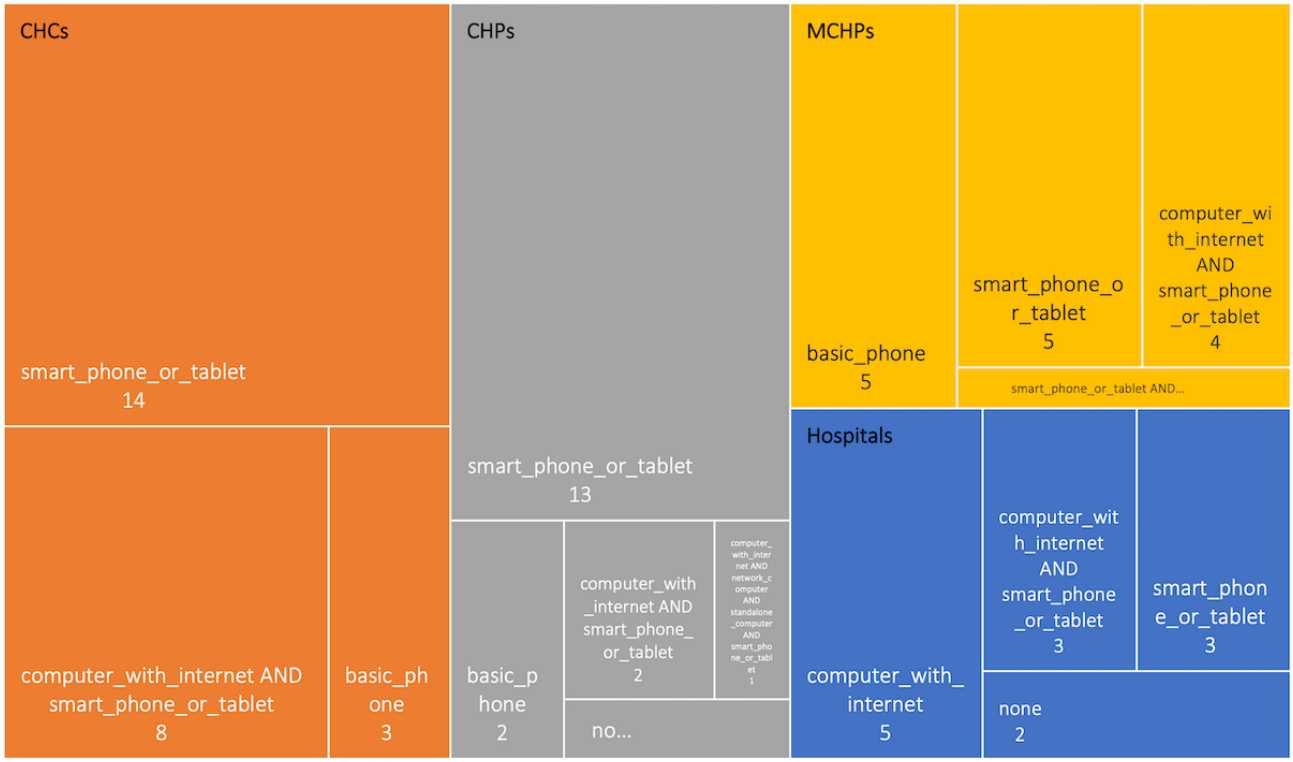


Table 3. Purpose of the digital health services and applications (health facility survey).

Health facility type and purpose of the services and applications	Number of health facilities				
	SA_1	SA_2	SA_3	SA_4	SA_5
Hospital					
data_services	5	0	0	1	1
healthcare_provider data_services	4	3	0	0	0
client data_services	1	1	0	0	0
client healthcare_provider, health_systems_administrator, and data_services	1	0	0	0	0
health_systems_administrator	1	0	0	0	0
healthcare_provider, health_systems_administrator, and data_services	1	1	0	0	0
health_systems_administrator and data_services	0	0	2	0	0
Client	0	0	1	0	0
Community health centers					
healthcare_provider and data_services	8	1	0	0	0
healthcare_provider, health_systems_administrator, and data_services	8	4	0	0	0
data_services	4	5	0	0	0
Client and data_services	2	0	0	0	0
Client, healthcare_provider, health_systems_administrator, and data_services	1	2	1	0	0
health_systems_administrator and data_services	1	1	0	0	0
healthcare_provider	1	0	0	0	0
Client and healthcare_provider	0	0	0	1	0
Community health posts					
healthcare_provider and data_services	10	2	0	1	0
data_services	2	0	0	0	0
Client	1	1	0	0	0
Client and data_services	1	1	0	0	0
Client, healthcare_provider, and data_services	1	1	0	0	0
Client, healthcare_provider, health_systems_administrator, and data_services	1	0	1	0	0
Client and health_systems_administrator	1	0	0	0	0
health_systems_administrator	1	0	0	0	0
health_systems_administrator and data_services	1	1	0	0	0
healthcare_provider	1	0	0	0	0
Maternal and child health posts					
data_services	4	2	1	0	0
healthcare_provider and data_services	4	0	0	0	0
Client and data_services	2	1	0	0	0
healthcare_provider, health_systems_administrator, and data_services	2	2	0	0	0
health_systems_administrator	1	1	1	0	0
health_systems_administrator and data_services	1	1	0	0	0
healthcare_provider and health_systems_administrator	1	0	0	0	0
Client, healthcare_provider, health_systems_administrator, and data_services	1	0	0	0	0

Figure 3. Access techniques used for digital health services and applications (health facility survey). CHC: community health center; CHP: community health post; MCHP: maternal and child health post.



Implementing Partner Survey Finding

Implementing partners used the services and applications shown in the word-art in Multimedia Appendix 1. The majority of the tools used were for data collection, processing, and reporting. The majority of the implementing partners supported the use of the DHIS, either through the national instance or a different instance. All surveyed implementing organizations noted that the status of their digital health effort was "active and working."

Information Sharing

The findings from all the 13 districts show that all DHMTs share both service delivery and implementation information with health facilities in their district. All but one share (or send) both implementation and service delivery data with the central MoHS, nongovernment organizations (NGOs) or other implementing organizations. In total, 6 districts shared this information via email, 6 shared it in print format, and 3 reported sharing data via SMS. All districts submit data to the NHMIS web portal in the required format every month. Health facility information sharing is along the aggregate data reporting only to the DHMT or supporting NGOs. At PHUs, 78% share aggregate data with the DHMT only, and the other 22% share the aggregate data with the DHMT and NGOs. Similarly, hospitals share aggregated service delivery and implementation data with 16% either sharing to NGO or to no one as shown in Table 4. No health facility or district shares individualized information between different applications.

The majority of the implementing partners reported having a written standard operating procedure to facilitate data exchange at the health facilities they supported. Almost all partners surveyed shared data in a government-approved format, in addition to other formats. The majority of the partners shared data with health facilities, DHMTs, and the MoHS.

Table 4. Distribution of where aggregate data are sent by hospitals (health facility survey).

Where the data are sent	Proportion, %
To no one	8
District Health Management Team	23
Central Directorate of Policy, Planning and Information	23
District Health Management Team and central Directorate of Policy, Planning and Information	23
District Health Management System, nongovernment organizations, and central Directorate of Policy, Planning and Information	15
Nongovernment organizations	8

Data Use

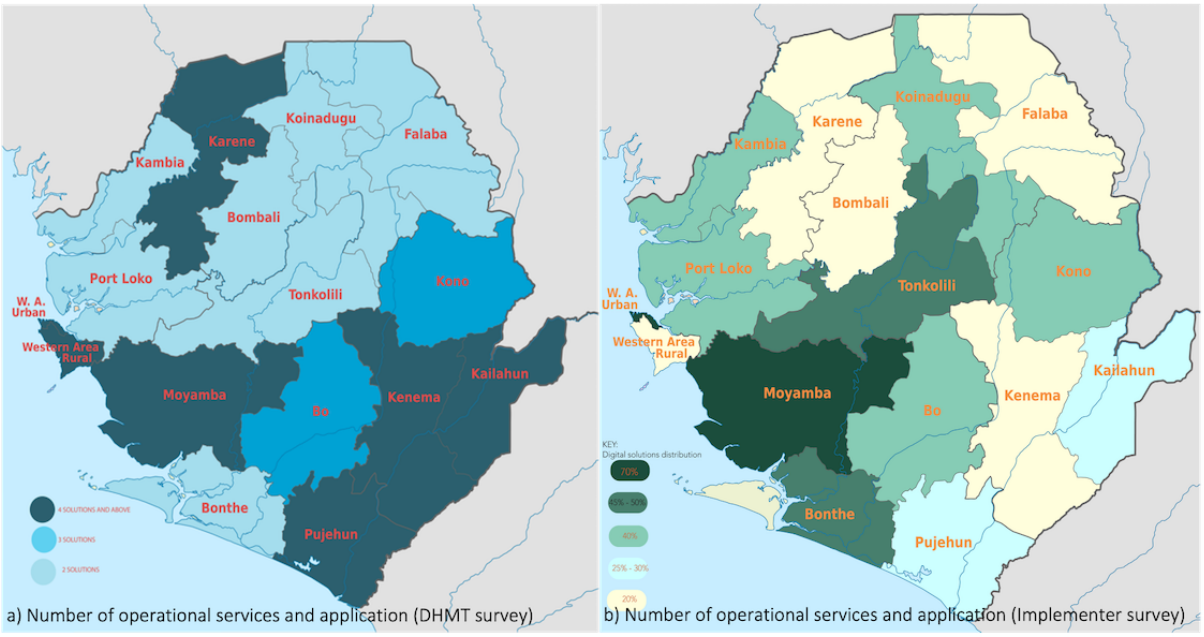
All surveyed health facilities reported using data for decision-making.

Discussion

Solutions (Services and Application)

Using the three structured interview approaches to determine the status of digital health solutions deployed in Sierra Leone highlighted the different dimensions and the distribution of these solutions. As shown in Figure 4a, the DHMT survey shows an even distribution of these tools across the districts. However, the implementing partner survey shows that the distribution is skewed to 2 districts, as shown in Figure 4b. Another dimension of this survey from the health facilities shows that most PHUs have 2 or more digital health services and applications. Furthermore, the majority of these solutions are intended for data services and use tablets as access mechanisms. On the other hand, hospital solutions use computers alone or computers and tablets. This is expected as the workload at hospitals often requires bigger-form factor hardware devices. Furthermore, hospitals have better electricity and internet-network infrastructure to support using computer-based services and applications (as against tablet-based services).

Figure 4. Number of operational services and application per district. (a) Number of operational services and applications (District Health Management Team survey); (b) number of operational services and applications (implementer survey).



Interoperability

Although stakeholders share information within and across institutions in government data sets, the structure and format of these data vary greatly (email, SMS, paper forms, and portal reporting). The implication of our findings is that no digital health foundational registry (patient, provider, practitioner, and health terminology classification) is used by any of these tools. There should be coordination around a standardized data format to reduce duplication among implementing partners, especially for individualized data-based solutions. Data-intensive digital health solutions should improve the feedback loop and data use, especially at health facility levels [13]. License-free individualized data sharing standards such as the Fast Healthcare Interoperability Resource should be explored [14]. The WHO's International Classification of Diseases or Systematize Nomenclature of Medicine Clinical Terms will be invaluable in designing an interoperability terminology interface as proposed in the digital health strategy [15]. In order to mitigate data blocking as classified by the US Health Insurance Portability and Accountability Act (HIPAA) [16], blockchain, an emerging technology that allows shared ownership and administration of data, can be used [17]. Data blocking has been responsible for limited interoperability, and low- and middle-income countries often use western regulations such as the HIPAA as the best practices benchmark. An insufficient feedback loop was identified by the surveys, especially at health facilities.

Big Data Opportunities and Digital Health Vision

Application of big data in health care can be critical, and evidence from around the world supports this [18]. However, different surveys of the application of big data in health care show that these applications use individual-level data rather than aggregate-level information [18]. Given that no solution currently shares or exchanges individual-level data (longitudinal patient data), opportunities for using facility-generated big data at present are greatly limited to aggregate disease surveillance only. Steps necessary to improve individualized information sharing are critical to achieving the vision of digital health. This can also impact the quality of health system data and the ability to use the data. Investment in a multi-sourced data triangulation system will be a low-hanging fruit for data interoperability and use especially at district levels.

Limitations

Given that the digital ecosystem is evolving, and owing to the rapid deployment of digital interventions fueled by the pandemic, we acknowledge that new solutions may have been deployed after the survey. However, this snapshot mapping will prove invaluable to policy makers. In addition, other key enabling environment components such as digital health infrastructure, workforce capacity, and funding remain key barriers to achieving these ideals.

Conclusions

This mapping from frontline health workers, policy makers, and implementers has shown that there are many digital health solutions in operation at health facilities in Sierra Leone. This study also shows that only aggregate-level data are shared for reporting and monitoring purposes only. Individualized information (or longitudinal patient data) is not currently processed for exchange among different solution providers. Hospitals mostly use computer-based solutions, while PHUs mostly use tablet-based solutions. No foundational digital health registry is used by any of the surveyed and mapped digital health solutions.

There are opportunities to leverage the 6 Vs of big data (value, volume, velocity, veracity, variability, and variety) to achieve the national digital health vision. Integrated care resulting from big data–facilitated electronic health records is only possible through individualized data-enabled care coordination [19].

Acknowledgments

The Ministry of Health and Sanitation was supported for this study by UNICEF with support from United States Agency for International Development. The University of Malta also supported this research. The authors would like to appreciate all survey respondents including district health medical officers, health facility officers in charge, Focus1000, eHealthAfrica, Emergency Operation Center, World Vision, GIZ, World Health Organization, UNICEF, Johns Hopkins University-Center for Communications Program, United Nations Population Fund, Cell, Care International, GOAL, Partners In Health, United Nations, Hellen Keller International, World Hope International, Community Health Workers Hub, and Monitoring and Evaluation, Directorate of Policy, Planning and Information.

Authors' Contributions

FS is the approving authority for this study. EC designed and prepared the initial draft of this paper. EF and AK facilitated the recruitment, training, field visits, and report reviews of the data collectors. RW and LG provided technical inputs and reviewed the manuscript draft.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Names of digital health services and applications (implementers' survey).
[PNG File , 52 KB - formative_v6i6e29930_app1.png]

References

1. Report of the Sierra Leone National eHealth Coordination Hub Launch. 2017. URL: https://mohs2017.files.wordpress.com/2017/12/report-of-the-ehealth-coordination-hub-launch_final.pdf [accessed 2021-01-20]
2. National Digital Health Strategy. Ministry of Health and Sanitation. 2018. URL: https://mohs2017.files.wordpress.com/2019/02/sl-national-digital-health-strategy-nov-2018.pdf [accessed 2022-04-28]
3. Directorate of Science Technology & Innovation. 2019. URL: https://www.dsti.gov.sl/wp-content/uploads/2019/11/Sierra-Leone-National-Innovation-and-Digital-Strategy.pdf [accessed 2021-01-21]
4. Andreu-Perez J, Poon CCY, Merrifield RD, Wong STC, Yang G. Big data for health. IEEE J Biomed Health Inform 2015 Jul;19(4):1193-1208. [doi: 10.1109/JBHI.2015.2450362] [Medline: 26173222]
5. Public Expenditure: Statistical Analyses 2020. HM Treasury. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/901406/CCS207_CCS0620768248-001_PESA_ARA_Complete_E-Laying__002_.pdf [accessed 2022-04-28]
6. Center for Disease Control. 2018 US Health Expenditure. Health Expenditure. 2018. URL: https://www.cdc.gov/nchs/fastats/health-expenditures.htm [accessed 2021-01-21]
7. World Health Statistics. World Health Organization. URL: http://www.who.int/gho/publications/world_health_statistics/2018/en/ [accessed 2022-04-28]
8. Transforming Health Care Through Big Data: Strategies for leveraging big data in the health care industry. Institute for Health Technology Transformation. URL: http://c4fd63cb482ce6861463-bc6183f1c18e748a49b87a25911a0555.r93.cf2.rackcdn.com/iHT2_BigData_2013.pdf [accessed 2022-04-28]
9. Ministry of Health and Sanitation. URL: https://mohs.gov.sl [accessed 2021-01-21]

10. dhis2. URL: http://www.dhis2.org/ [accessed 2020-08-28]
11. Design Powerful Mobile and Web Apps. CommCare. URL: http://www.commcarehq.org/home/ [accessed 2020-08-17]
12. Classification of digital health interventions v1.0: a shared language to describe the uses of digital technology for health. World Health Organization. URL: https://apps.who.int/iris/handle/10665/260480 [accessed 2021-01-20]
13. Hosseini M, Dixon BE. Syntactic Interoperability and the Role of Standards. In: Health Information Exchange: Navigating and Managing a Network of Health Information Systems. Cambridge, MA: Academic Press; 2016:123-136.
14. Operation-patient-match. HL7 FHIR Release 4. URL: https://www.hl7.org/fhir/operation-patient-match.html [accessed 2022-09-28]
15. International Statistical Classification of Diseases and Related Health Problems 10th Revision. URL: https://icd.who.int/browse10/2019/en [accessed 2021-01-10]
16. Zhang Y. Health Insurance Portability and Accountability Act (HIPAA). In: Encyclopedia of global health (volume 1). Thousand Oaks, CA: Sage Publications; 2008:806.
17. Chukwu E, Garg L. A Systematic Review of Blockchain in Healthcare: Frameworks, Prototypes, and Implementations. IEEE Access 2020;8:21196-21214. [doi: 10.1109/access.2020.2969881]
18. Khoury MJ, Ioannidis JPA. Medicine. Big data meets public health. Science 2014 Nov 28;346(6213):1054-1055 [FREE Full text] [doi: 10.1126/science.aaa2709] [Medline: 25430753]
19. The conceptual framework of interoperable electronic health record and ePrescribing systems. European Commission. 2008. URL: http://www.ehr-impact.eu/downloads/documents/EHRI_D1_2_Conceptual_framework_v1_0.pdf [accessed 2022-04-28]

Abbreviations

AI: artificial intelligence
DHIS2: District Health Information Software
DHMS: District Health Management System
DHMT: District Health Management Team
DMO: district health medical officer
DPPI: Directorate of Policy, Planning and Information
eIDSR: electronic Integrated Disease Surveillance and Response
HIS: health information system
MoHS: Ministry of Health and Sanitation
NGO: nongovernment organization
NHMIS: National Health Management Information System
PHU: primary health care units

Edited by E Meinert; submitted 26.04.21; peer-reviewed by J Ye, P Le; comments to author 09.06.21; revised version received 17.07.21; accepted 09.04.22; published 10.06.22.

Please cite as:
Chukwu E, Garg L, Foday E, Konomanyi A, Wright R, Smart F
Digital Health Solutions and State of Interoperability: Landscape Analysis of Sierra Leone
JMIR Form Res 2022;6(6):e29930
URL: https://formative.jmir.org/2022/6/e29930
doi:10.2196/29930
PMID:35687406

Original Paper

Patient and Health Professional Perceptions of Telemonitoring for Hypertension Management: Qualitative Study

Juliana Baratta[1*], MS; Cati Brown-Johnson[1*], PhD; Nadia Safaeinili[1], MPH; Lisa Goldman Rosas[1], PhD; Latha Palaniappan[1], MS, MD; Marcy Winget[1], PhD; Megan Mahoney[1], MD

Division of Primary Care and Population Health, Stanford School of Medicine, Palo Alto, CA, United States

[*]these authors contributed equally

Corresponding Author:
Juliana Baratta, MS
Division of Primary Care and Population Health
Stanford School of Medicine
3180 Porter Drive
Palo Alto, CA, 94304
United States
Phone: 1 5858807040
Email: jbaratta@alumni.stanford.edu

Abstract

Background: Hypertension is the most prevalent and important risk factor for cardiovascular disease, affecting nearly 50% of the US adult population; however, only 30% of these patients achieve controlled blood pressure (BP). Incorporating strategies into primary care that take into consideration individual patient needs, such as remote BP monitoring, may improve hypertension management.

Objective: From March 2018 to December 2018, Stanford implemented a precision health pilot called Humanwide, which aimed to leverage high-technology and high-touch medicine to tailor individualized care for conditions such as hypertension. We examined multi-stakeholder perceptions of hypertension management in Humanwide to evaluate the program's acceptability, appropriateness, feasibility, and sustainability.

Methods: We conducted semistructured interviews with 16 patients and 15 health professionals to assess their experiences with hypertension management in Humanwide. We transcribed and analyzed the interviews using a hybrid approach of inductive and deductive analysis to identify common themes around hypertension management and consensus methods to ensure reliability and validity.

Results: A total of 63% (10/16) of the patients and 40% (6/15) of the health professionals mentioned hypertension in the context of Humanwide. These participants reported that remote BP monitoring improved motivation, BP control, and overall clinic efficiency. The health professionals discussed feasibility challenges, including the time needed to analyze BP data and provide individualized feedback, integration of BP data, technological difficulties with the BP cuff, and decreased patient use of remote BP monitoring over time.

Conclusions: Remote BP monitoring for hypertension management in Humanwide was acceptable to patients and health professionals and appropriate for care. Important challenges need to be addressed to improve the feasibility and sustainability of this approach by leveraging team-based care, engaging patients to sustain remote BP monitoring, standardizing electronic medical record integration of BP measurements, and finding more user-friendly BP cuffs.

(JMIR Form Res 2022;6(6):e32874) doi:10.2196/32874

KEYWORDS

hypertension; remote blood pressure monitoring; precision health; mobile phone

Introduction

Background

Hypertension is the most prevalent and important risk factor for cardiovascular disease, affecting 1 in 4 adults worldwide [1]. According to the Centers for Disease Control and Prevention, in 2014, hypertension was the underlying cause of death of 410,000 Americans, with >1100 deaths each day [2]. Nearly half of US adults are hypertensive, with a blood pressure (BP) >130/80 mm Hg. If left untreated, hypertension increases the risk of heart attack, stroke, kidney disease, and Alzheimer disease [3]. The risk of these adverse consequences can be mitigated through BP reduction by adhering to hypertension treatment, including behavior modification (eg, low-sodium diet and regular exercise) and taking medication, as recommended by the American Heart Association (AHA) guidelines [2].

Adherence to hypertension treatment has been associated with reductions of 35% to 40% in stroke incidence, 20% to 25% in myocardial infarction incidence, and 50% in heart failure incidence [4]. However, only 54% of US patients with hypertension have controlled BP [5]. Challenges to achieving controlled BP include failure to respond to medication, treatment side effects leading to subpar adherence, and lack of engagement in preventive behaviors such as adopting a healthy diet and increasing physical activity [6,7]. Thus, there is a need to advance hypertension management through individualized approaches that engage patients.

Precision health is an emerging approach to patient-centered care [8] that incorporates patients' variations in genes, environment, and behavioral lifestyle to construct personalized treatment and prevention approaches [9]. From a population health perspective, precision health can improve prevention of heart disease by defining subgroups of patients with hypertension that may benefit from specific therapies [6]. For the individual patient, hypertension management using precision health can enable better targeting of personalized treatment options by identifying high-risk patients or those in early disease stages in the hopes of averting negative outcomes in the future [10]. Previous research has shown that remote BP monitoring combined with health coaching [11] or pharmacist management [12] improves BP control by providing consistent and accurate BP data to both patients and physicians, which is then used to inform the selection of more effective treatment options [13]. Remote BP monitoring has also been shown to empower patients in relation to managing their hypertension [14] and improve the patient-clinician alliance [15].

As a result of the COVID-19 pandemic, there is an even greater clinical need for remote BP monitoring for hypertension management, as reflected by recent policies. On March 20, 2020, the Food and Drug Administration issued an enforcement policy for the expedited use and availability of digital remote monitoring equipment to facilitate patient monitoring during COVID-19 [16]. Although this policy is only to remain in effect during the pandemic, recent uptakes of remote BP monitoring may be sustained owing to support from other policies. For example, in 2019, the National Committee for Quality Assurance updated the hypertension quality measure to allow BP readings to be taken using remote patient monitoring devices and telehealth encounters to satisfy certain components of the quality measure [17]. The use of home BP measurement is also recommended for the ongoing diagnosis and treatment of hypertension in both the 2020 International Society of Hypertension Global Hypertension Practice Guidelines [18] and the 2017 American College of Cardiology and AHA Blood Pressure Guidelines [19].

As a result, remote BP monitoring is becoming increasingly important for hypertension management and, more broadly, in population health programs. Despite recommendations, adoption is precluded by mediators and moderators of remote BP monitoring integration, including the usability of the digital health tools, ease of clinical workflow incorporation, and availability of technical support [20]. Research is still needed to identify the best practices to sustain remote BP monitoring by overcoming barriers on both the patient side (ie, reductions in motivation) and the clinic side, such as the high costs of hardware maintenance and related software for digital health monitoring [21].

Objectives

Stanford conducted a precision health pilot, Humanwide, to assess the feasibility of embedding a precision health model in a community-based primary care clinic [10]. The goal of Humanwide was to deliver precision health through a combination of "high tech and high touch" care via the use of genetic and pharmacogenomic testing, digital health monitoring, and intensive one-on-one health coaching in the context of team-based primary care [22]. To our knowledge, this is the first implementation of a multipronged precision health delivery model integrated into a primary care clinic. Considering the continuing implementation hurdles that BP management poses and the potential of precision health in this space of patient care, the purpose of this analysis was to formally assess the implementation outcomes of *feasibility*, *acceptability*, *appropriateness*, and *sustainability* [23] of hypertension management via Humanwide and examine multi-stakeholder perceptions of this approach.

Methods

Overview

The implementation of Humanwide took place between March 2018 and December 2018. Patient participation in this study was entirely optional. After enrollment in the pilot study, patient information was shared securely with researchers (NS and CBJ), who then contacted patients to determine their interest in participating in the evaluation. Before conducting the interviews, the evaluation team obtained the participants' verbal consent. The participants were made aware that the interviews would be confidential and in no way affect their care. Interviews and transcripts were only accessed by the external evaluation team and not by Humanwide health professionals. Audio and transcription files were maintained in Health Insurance Portability and Accountability Act–compliant box files. The deidentified aggregate findings were shared with health professionals and Humanwide team members.

Overview of Humanwide Pilot

Humanwide entailed four components that were added to standard primary care: (1) a baseline wellness visit to assess patient lifestyle, demographics, and socioenvironmental health factors, with follow-up health coaching visits as needed; (2) digital health remote biometric monitoring through the HealthKit app (Apple Inc) [24], including Bluetooth-enabled home scale glucometer, BP cuff for intermittent remote BP readings via the Withings device and app [25], and pedometer; (3) family history assessment and follow-up genetic testing for patients identified from the assessment as at risk for breast cancer, familial hypercholesterolemia, or Lynch syndrome; and (4) pharmacogenomic testing to examine a patient's likely response to a given class of drugs given their genetic makeup [10]. Implementing these components required care coordination across multiple physicians and specialists, medical assistants (MAs), a pharmacist, a behavioral health practitioner, a registered nurse, and a genetics counselor. The inclusion criteria were (1) adults aged >18 years, (2) seeing a health professional at the pilot study clinic, (3) having a smartphone (to take part in the digital health component), and (4) having time to participate. Health professionals recruited patients that they felt could benefit from Humanwide components, such as medically complex patients managing more than one chronic illness. Attention was given to recruiting diverse patients with respect to age, race and ethnicity, gender, and medical complexities. Primary care health professionals invited 69 patients to participate in Humanwide, and 50 (72%) enrolled. The 19 (28%) patients who declined enrollment reported lack of time as their main reason for nonenrollment.

Interviews and Data Collection

All patients and health professionals involved with Humanwide were eligible for interviews, and patient recruitment for interviews occurred simultaneously during Humanwide enrollment. Health professionals in the primary clinic and specialists outside of the clinic who were contributing to patient care as part of the pilot were included in the interviews. Patients and health professionals were recruited for interviews via convenience sampling. We tracked roles (ie, primary care physician [PCP], nurse, or pharmacist) to aim for a purposive sample.

We developed a semistructured interview guide to assess the implementation outcomes of *feasibility*, *appropriateness*, and *acceptability* of Humanwide [23]. Interview questions assessed perceptions of each pilot component—genetic testing, pharmacogenomics, digital health, and health coaching—and addressed recommendations for future implementation of Humanwide with respect to hypertension management. As the interviews were semistructured, if a patient mentioned hypertension or remote BP monitoring, the questions were then phrased to assess how the pilot components affected their hypertension management (see Table 1 for a summary of the interview guide questions pertaining to hypertension). A total of 3 researchers trained in qualitative methods (NS, JB, and CBJ) conducted audio-recorded interviews in person in a private conference room or over the phone. To the researchers' knowledge, no other individuals were present during the interviews. Patient interviews ranged from 17 to 36 minutes (mean 25, SD 6.7 minutes). Health professional interviews ranged from 22 to 60 minutes (mean 45, SD 13.9 minutes). No financial or other compensation was provided for participating in the interviews.

Table 1. Summary interview guide for health professionals.

Category	Question
• Precision health and hypertension	• What was your experience like with hypertension management in Humanwide?
• Remote blood pressure monitoring	• Which parts of remote blood pressure monitoring worked well? Which parts did not work well? • What kind of expectations did receiving blood pressure data place on you?
• Genetic testing and pharmacogenomics	• Can you tell us if and how genetic or pharmacogenomic testing affected hypertension management?
• Health coaching	• Can you tell us if and how one-on-one health coaching affected hypertension management?
• Implementation feasibility and sustainability	• What are your thoughts about the sustainability of hypertension management in Humanwide? What would make this approach more sustainable?

Data Analysis

We used a hybrid qualitative approach integrating a priori and emergent themes [26]. A priori subjects of interest included hypertension, digital health, pharmacogenomics, genetic testing, and health coaching. Emergent themes were identified via thematic analysis, which involved careful reading and rereading of the transcripts in line with the inductive approach [26]. The analysis involved 3 steps. First, JB read all patient and health professional transcripts using NVivo 11 software (QSR International) and coded them using the a priori subjects of interest [27]. Second, JB extracted emergent themes from all transcripts with input from the full authorship team. Third, emergent themes were assessed in relation to the following implementation science outcomes based on Proctor et al [23]: *acceptability* (satisfaction with various aspects), *appropriateness* (perceived fit), *feasibility* (suitability for everyday use and ability to be carried out considering resources, training, and staff), and *sustainability* (facilitators and barriers to spread). Coding

questions and novel emergent codes were discussed during weekly meetings with CBJ, NS, and JB over the course of 4 months. In total, 2 researchers (CBJ and NS) conducted quality checks and verified a final coding schema that included codes for a priori concepts and constructs as well as emergent themes.

Ethics Approval

This study was given a nonresearch determination and Human Subjects Research Exemption protocol 43279 by the Stanford Institutional Review Board.

Results

Participants

Of the 50 patients in Humanwide, 16 (32%) participated in the qualitative evaluation. The interviewed patients were diverse—50% (8/16) were non-White, and 56% (9/16) were women (Table 2)—and representative of the 50 patients enrolled based on race and ethnicity, gender, and age. Patients who explicitly referenced hypertension, remote BP monitoring, or both in their interviews were included in the analysis (10/16, 63%). We interviewed 11 health professionals in the Humanwide pilot clinic, including 9 (82%) PCPs, 1 (9%) pharmacist, and 1 (9%) registered nurse; interviews referencing hypertension or BP monitoring were included in the analysis (6/11, 55%). We also interviewed 4 key informant specialist medical doctors (MDs) involved in the pilot whose practices were outside the primary care clinic. The key informants were a cardiovascular geneticist, a pharmacogenomic specialist, a physician expert in biomedical informatics, and a physician specializing in the management of chronic medical conditions, particularly hypertension (Table 3).

Table 2. Patient characteristics (N=16).

Characteristic	Patients, n (%)
Age (years)	
30 to 39	3 (19)
40 to 49	6 (38)
50 to 59	5 (31)
60 to 69	2 (13)
Race and ethnicity	
White	8 (50)
Asian	5 (31)
Other	3 (19)
Gender	
Men	7 (44)
Women	9 (56)

Table 3. Health professional characteristics (N=10).

Characteristic	Health professionals, n (%)
Profession	
Primary care physician	4 (40)
Specialist medical doctor	4 (40)
Pharmacist	1 (10)
Nurse	1 (10)
Gender	
Women	7 (70)
Men	3 (30)

Emergent Qualitative Themes

Overall, the remote BP monitoring component of Humanwide was the only component that the patients mentioned as contributing to hypertension management. Other components (ie, health coaching, pharmacogenomics, and genetic testing) were not mentioned in conjunction with hypertension management. The participants reported that remote BP monitoring led to mixed increases in patient motivation, enhanced patient-clinician engagement, and improved patient hypertensive management. The participants discussed varied efficiency with remote BP monitoring, and the health professionals were overwhelmed by unfiltered BP data and by providing individualized feedback. The health professionals proposed solutions to these barriers, including managing data through electronic medical record (EMR) settings and leveraging

team-based care. Table 4 summarizes the themes that emerged from the analysis along with illustrative quotes from the interviews mapped to implementation outcomes.

Table 4. Themes and illustrative examples mapped to implementation outcomes.

Implementation outcome and emergent theme	Illustrative examples
Acceptability	
Increased patient motivation	"I figure if I am measuring my blood pressure on a regular basis and I am noticing that it is high, it should be a mental kick in the head that says 'Oh hey, I am going to do something about this.'" [Patient 8]
Increased patient-clinician engagement	"I think the key thing that it does is it builds that relationship. If I have to see you in three months, you're not going to think about doing your blood pressure. Maintaining, pushing a little more of, 'How are you doing? What's going on?' Just trying to understand that. I think our goal was to try to be more engaged in their lives outside of clinic." [Health professional 4, PCP[a]]
Appropriateness	
Remote blood pressure monitoring perceivably improves patient hypertensive management	"And also some days I feel like really dizzy and like I am about to faint so then I immediately check my blood pressure to see if it is normal because sometimes I feel like my heart is pounding very fast, so anything that I feel that is not right or something is different, I immediately check to see my blood pressure." [Patient 5]
Efficiency with remote blood pressure monitoring	"Just because we were able to access how things are at home. We could see what their blood pressure looks like at home versus in the clinic. It saves a lot of time when you're meeting with patients cause you have all that information ahead of time." [Health professional 11, specialist MD[b]]
Feasibility	
Efficiency with remote blood pressure monitoring	"Then I can show them the data, hey, look my blood pressure monitor taken this day is this, taken this day is this. Yeah I like the application itself. The application piece is good. It keeps my historical data." [Patient 7]
Technical difficulties with blood pressure cuff	"You have to make sure your bluetooth is on and then you have to make sure everything pairs and sometimes with the blood pressure cuff it will like go through the whole thing where it is squeezing and whatever and then it will be like, oh, error, it did not read." [Patient 6]
Time lost	"But it does put that added burden back on me to look through it [BP[c] readings]. I'm getting five to ten trackers, tracking notices, now every single day. The patients have those tracking information back for me. But if I'm trying to look at everything, which...that's my goal, then it's too much." [Health professional 1, PCP]
Sustainability	
Managing data through EMR[d] settings	"EPIC has some tools to visualize data in general and it is incorporated in those same views. Just as a normal PC doc would visualize BP data um same basic kind of mechanisms and dashboards. But I think it is an area of active discussion and debate. Like are the tools for health professionals and patients to interact with their data are they as good as they could be and how can we make them more useful." [Health professional 13, specialist MD]
Sustaining patient motivation	"I think we're trying with the digital health, and I think we'll continue to try. I think, again, it has huge benefit for the people that will do it. I think we need to figure out how to get people to do it, but I think it still has great potential." [Health professional 5, PCP]
Need for team-based care	"I think our goal was to try to be more engaged in their lives outside of clinic. That was one of the purposes of these things. How do we do that? Do we need more support doing that? Could an MA[e] do that?" [Health professional 2, PCP]

[a]PCP: primary care physician.

[b]MD: medical doctor.

[c]BP: blood pressure.

[d]EMR: electronic medical record.

[e]MA: medical assistant.

Implementation Outcomes

Overall, remote BP monitoring was acceptable and appropriate for hypertension management, whereas other components (ie, health coaching, pharmacogenomics, and genetic testing) were not mentioned as contributing to hypertension care. The patients and health professionals reported some barriers to feasibility and sustainability but recommended solutions to overcome these concerns for future implementation.

Acceptability

Overview

The use of remote BP monitoring to facilitate hypertension management was acceptable to most patients (13/16, 81%) and health professionals (8/10, 80%) and led to increased patient-health professional engagement; however, acceptance was mixed as not all patients with hypertension used it. On the positive side, the patients enjoyed receiving individualized feedback and treatment guidance from physicians; in some cases, this improved the patients' motivation to make lifestyle modifications and engage in remote BP monitoring. In contrast, there were several technological glitches with the wireless Bluetooth BP cuff used, and it was thus not deemed an acceptable device in the long term by patients or health professionals.

Mixed Patient Motivation

The patients perceived remote BP monitoring with the wireless Bluetooth cuff as one of the main contributors to motivation, although maintenance was difficult for some (3/16, 19%). Patients with hypertension reported an improved desire to make behavioral changes such as implementing a routine exercise regimen. They described how seeing their BP measurements made them more conscientious of their health, which increased their motivation for self-management. The patients also reported an increased sense of accountability to their health professional, which further contributed to motivation:

> *I mean the attention from the medical staff and the fact that it is there in the app [on my smart phone] is going to make me become more efficient at [blood pressure monitoring].* [Patient 7]

The patients perceived health professional feedback as the "cue to take action" (patient 3) to adhere to treatment.

Despite increased patient motivation, the health professionals reported that patients did not sustain remote BP monitoring:

> *The [challenge] of Humanwide is that it probably requires some thought on how to adequately coach patients so that they feel really engaged in the use of wearables.* [Health professional 12, specialist MD]

A total of 13% (2/15) of the health professionals discussed that, as patients with hypertension controlled their BP, they sometimes stopped using the cuff because they felt less incentivized. In addition, the health professionals mentioned that some patients who needed remote BP monitoring did not participate:

> *It's insightful to me that it's just really hard to get buy-in for that [blood pressure cuff]...For those that are interested, it's great, but it's very few.* [Health professional 2, PCP]

A few (3/15, 20%) health professionals discussed not having the bandwidth to reach out to all patients with hypertension who were less engaged in remote BP monitoring; instead, the health professionals capitalized on patients who were engaged using a "nudge-based approach" (health professional 12), "frequent touches" (health professional 4), and "targeted management plans" (health professional 6).

Appropriateness

Overview

The health professionals and patients perceived remote BP monitoring to be appropriate based on their experiences of improved hypertension management and patient-health professional engagement, with fewer clinic visits. The health professionals reported that diagnosing patients with masked hypertension with the help of remote BP monitoring allowed patients to incorporate lifestyle changes earlier to achieve BP control. In addition, the health professionals reported titrating the medications of patients with hypertension faster because of remote BP monitoring, potentially leading to moderate improvements in overall clinic efficiency. Most patients with hypertension (15/16, 94%) enjoyed the digital health component and found it helpful to achieve controlled BP because they became more aware of their BP and knowledgeable of how to take part in their treatment.

Increased Patient-Clinician Engagement

The use of remote BP monitoring appeared to promote patient-clinician engagement and treatment adherence. The health professionals kept in touch with patients by providing individualized feedback on BP data. A patient mentioned the following:

> *My doctor called me and noted that my blood pressure was higher than it should be and let me know about that, that I need to take action on it.* [Patient 8]

The patients with hypertension noted that they really "enjoyed" (patient 3) and "appreciated" (patient 6) receiving feedback from the clinical care team. In addition, the patients perceived this feedback as helpful for achieving their behavioral goals.

The health professionals similarly implied that the increased connection with their patients improved patient understanding of and adherence to their treatment regimen:

> *We have a tighter relationship. When I reach out, it's not like I saw them last year. They'll maybe listen a little bit more. We'll have a little bit more of a conversation.* [Health professional 10, pharmacist]

Most health professionals (12/15, 80%) mentioned that receiving the BP data placed an expectation on them to reach out to their patients to provide feedback.

Improved Patient Hypertensive Management

The participants perceived improved hypertensive management as a result of remote BP monitoring. The participants reported that remote BP monitoring improved "awareness of what was going on at home" (patient 3), as best described by a patient:

> *Now that there is an application and a digital record that allows me to be more conscientious, "Oh wow, it has been a week, or it has been 4 days, or it has been 3 days," and, you know, that is now in my mind being more attentive about checking my blood pressure.* [Patient 7]

The health professionals mentioned at-home BP data serving as a "checkpoint" (health professional 11) to ensure that patients were well-managed. A total of 20% (3/15) of the health professionals noted that some patients had normal BP in the clinic but elevated BP out of the clinic:

> *Individuals who never had a diagnosis of hypertension were getting blood pressure measurements in the range that would meet the criteria for hypertension as a diagnosis, and then they were able to...make lifestyle changes that were very tailored to them, that then reduced their blood pressure, and prevented hypertension.* [Health professional 6, PCP]

Home BP monitoring was coupled with regular measurements in the clinic to help ensure accuracy and precision between home and clinic BP measurements.

Feasibility

Overview

There were concerns regarding the long-term feasibility of hypertension management in Humanwide because of the following barriers: (1) limitations of the BP cuff, including limited sizing, which may have led to inaccurate results for those with high BMI; (2) technical difficulties because of wireless connectivity issues; (3) time lost sifting through overwhelming amounts of BP measurements in the EMR; and (4) the number of health professionals needed to provide individualized feedback to patients. Unfortunately, the patients and health professionals became frustrated with the BP cuff's technological issues, which led to drops in engagement with remote BP monitoring throughout the pilot. The health professionals also reported significant time lost reading through patient BP data to identify clinically actionable measurements. Although the health professionals believed that providing individualized feedback was beneficial to patients, they felt it was not sustainable unless they used MAs or pharmacists to share this responsibility.

Mixed Efficiency With Remote BP Monitoring

There were mixed reports on the efficiency in managing patients with hypertension as a result of remote BP monitoring. Although some participants felt that remote BP monitoring made it easier to track BP and contributed to increased clinic efficiency, others mentioned technological glitches reducing efficiency and motivation. The patients reported that the wireless Bluetooth feature of the BP cuff saved time and made tracking BP data easy:

> *Yeah I like the application itself. The application piece is good. It keeps my historical data.* [Patient 1]

A health professional ratified the value of this wireless Bluetooth technology considering the BP reads auto-populate the patient's app and medical chart:

> *I don't have to call my patient asking them to read back some blood pressures, and conversely I know when they are not recording their blood pressures.* [Health professional 13, specialist MD]

This example illustrates the potential of remote BP monitoring to improve health professionals' awareness of patient progress and inform individualized treatment.

The health professionals also discussed that having patient BP data before a clinic visit saved time and made the visit more productive. The health professionals implied that it improved clinic efficiency as it took fewer visits to stabilize patients on their medication:

> *I think we're able to titrate them quicker, to make the blood pressure medicine changes quicker than we would having to come into the office every three to six months. So, we get them to goal quicker.* [Health professional 2, PCP]

Of the 15 health professionals, 9 (60%) corroborated that they became "pretty efficient in terms of the workflow" (health professional 4) since implementing remote BP monitoring.

Although the participants discussed increased efficiency with remote BP monitoring, many expressed concerns over technological glitches with the wireless Bluetooth BP cuff. The patients mentioned difficulties with the BP readings not syncing to their EMR and the cuff not taking measurements:

> *They are a little glitchy. The blood pressure cuff is probably the easiest [remote monitoring digital health tool] to use but there are several steps involved.* [Patient 6]

The health professionals echoed this sentiment, reporting that patients did not regularly monitor BP because of glitches with the technology:

> *They [patients] get frustrated when they put the cuff on, and it doesn't work. They don't want to do it.* [Health professional 2, PCP]

Technological difficulties also affected health professional workflow because the health professionals often had to help patients troubleshoot over the phone or schedule a patient visit:

> *With the BP cuff, sometimes there were technical issues and they [patients] kind of knew to ask me or I gave them resources to call.* [Health professional 11, specialist MD]

Indeed, the pilot intake process included a full hour with an MA entirely devoted to troubleshooting digital health devices. The patients mentioned needing to adjust their arm position for the cuff to work, and a patient required an additional office visit for guidance.

Sustainability

Overview

The health professionals mentioned several concerns related to the sustainability of remote BP monitoring as part of hypertension management. They discussed the negative impacts on the sustainability of remote BP workflows and patient

motivation and engagement from feasibility issues: BP cuff technical glitches reducing motivation for patients, health professional time lost sifting through BP measurements, and time constraints limiting health professionals' individualized feedback to patients. According to patient and health professional interviewees, the patients measured their BP daily, especially during the first half of the intervention period (March 2018 to July 2018), with a gradual drop in frequency of BP measurements in the latter half of the pilot.

The health professionals suggested several solutions to overcome feasibility and sustainability issues, including using a different wireless Bluetooth cuff in the future, providing patients with more information or guidance on how to properly use the cuff, better leveraging team-based care, and enabling measurement of BP data through automated EMR settings. Ideally, guidance for BP cuff use would be accessed outside the clinic through video tutorials or on-demand technological support. EMR settings could be adjusted such that health professionals receive an alert only when a patient's BP is above a certain cutoff. A final suggestion was to incorporate artificial intelligence (AI)-based BP cuffs that would only surface alarming BP measurements. Overall, a great deal of care coordination, technological improvements, and approaches to sustain remote BP monitoring needs to be addressed to achieve long-term sustainability.

Overwhelmed by Unfiltered Data and Providing Individualized Feedback

The health professionals received a deluge of BP data in the EMR from remote monitoring and reported time lost as a result of reviewing these data and following up with patients:

> *I have to respond to them or do something with the information. But it's now turning out to be five to ten every single day that I'm getting. And obviously I think it's useful, which is why I'm doing it. But at some point, I'm gonna say, "This is just too much."* [Health professional 2, PCP]

The health professionals reported that providing tailored feedback to patients was appropriate for hypertension management but acknowledged that the current workflow, where health professionals are solely responsible for communication, might not be acceptable in the long term. Conversely, several health professionals implied that daily communication with patients was worth their time even though there were no shortcuts:

> *You can't just say, "Do this," which is fine. I mean, you have to understand that it's just not going to be a quick [fix]. I mean, it's that relationship, right? You're building that relationship.* [Health professional 2, PCP]

A health professional even felt that daily communication was sustainable:

> *The frequent touches and the frequent follow-up is very doable.* [Health professional 4]

Health Professional Proposed Solutions

Managing Data Through EMR Settings

The health professionals managed BP data through EMR settings to reduce extra time spent sifting through BP measurements by setting "overs and unders" (health professional 12) so they would only be notified when a patient's BP was too high or low. Another strategy included setting the EMR to receive BP data every 2 weeks. The health professionals reported that it was not feasible to check BP measurements in real time. A health professional emphasized advising patients with hypertension to seek emergency help as they normally would if they encountered alarming BP measurements in conjunction with signs and symptoms of a hypertension emergency. If the patients were experiencing high BP measurements in isolation, they were advised to reach out to their PCP for potential medication adjustments or follow-up visits.

A few health professionals (3/15, 20%) also suggested developing graphical displays of BP measurements over time that could be visualized within the EMR to capture trends, outliers, and average BP values:

> *Graphical displays are the most meaningful for the health professional who is looking at it [BP data]. Most of my colleagues, when we are in clinic, we look at the graphical trend over time and we look at the average.* [Health professional 6, PCP]

Optimal visualization of BP data via the EMR remains an area of "active discussion and debate" (health professional 13).

Need for Team-Based Care

Many health professionals emphasized the importance of team-based care to enable successful hypertension management. A total of 40% (6/15) of the health professionals mentioned using MAs and pharmacists when treating patients with hypertension and the importance of everyone working at the top of their license. To best integrate remote BP monitoring, 13% (2/15) of the health professionals discussed needing to alter the typical patient-health professional model:

> *If you just apply technology to existing workflows and models you are not necessarily going to have better outcomes. You need to figure out, "What care models do these new technologies enable?" And it is things like centralization and different care team members interacting differently to data.* [Health professional 13]

The health professionals expressed that not all health care systems will have the resources to implement remote BP monitoring and individualized health professional feedback. However, a health professional discussed the potential for web-based patient management with the help of remote BP monitoring:

> *We are trying to build on top of that kind of precision health approach with a protocolized team-based strategy for remote patient monitoring and appropriate care referral, with the thought being that most physicians or health professionals may have difficulty seeing their patients more regularly than 3*

> *months just because access is limited.* [Health professional 12, specialist MD]

The health professionals also mentioned the ability of remote BP monitoring to reduce patient visits and, therefore, increase clinic efficiency.

Discussion

Principal Findings

This study explored emergent themes along with facilitators and barriers to hypertension management in the first reported precision health pilot study integrated into a primary care setting. The participants reported that remote BP monitoring led to improvements in patient treatment adherence and lifestyle behavior changes. These accounts are similar to those of previous studies showing improved medication management and adoption of lifestyle changes in patients using remote BP monitoring and health coaching [28,29]. The purpose of this evaluation was to assess the implementation outcomes for the Humanwide precision health pilot to inform future expansion of the intervention.

Most health professionals (9/15, 60%) stated that remote BP measurements are helpful for managing patients with hypertension as they are "actionable" and "are providing a new source of ground truth" (health professional 16, specialist MD). The AHA similarly recommends remote BP monitoring for a more comprehensive view of patients' BP control [30]. Other benefits of remote BP monitoring coupled with office BP measurements include better prediction of cardiovascular morbidity and mortality, improved patient understanding of hypertension management leading to better treatment adherence, and increased detection of BP variability [31].

The health professionals mentioned difficulty in obtaining patient buy-in to start remote BP monitoring. PCPs from a recent qualitative study suggested patient education sessions to enhance patient engagement by emphasizing how remote BP monitoring can be a source of individual empowerment in clinical care [32]. Another study proposed a model based on the business process management paradigm to empower patients by setting negotiated health goals and providing consistent lines of communication between the patient and their health care team to help trigger initial engagement in remote BP monitoring [33].

Along with the difficulty in encouraging patients to initially engage in remote BP monitoring, there were concerns that motivation was not sustainable in the long term. Some patients reportedly stopped using their cuff near the end of the pilot, potentially because they achieved BP control. However, patients who achieve well-controlled hypertension should continue monitoring their BP on a semiregular basis [30]. Common patient-level moderators of BP control include self-efficacy, self-awareness, and education [34]. Self-monitoring of lifestyle behaviors (eg, diet, physical activity, and sleep) tends to be an effective tool for *changing* behavior but may be less well-suited for *maintaining* behavior as it is challenging to do so in the long term [31]. Although some studies show that remote BP monitoring combined with telehealth counseling could improve adherence to hypertension care [35], most patients (14/16, 88%) did not report engaging in health coaching sessions. We believe that this lack of engagement was potentially due to greater interest in the more novel pilot components of genetic testing, pharmacogenomic testing, and Bluetooth devices. Although the participants did receive genetic and pharmacogenomic testing, the results of these tests did not significantly affect their hypertension management or treatment regime.

Tools to improve long-term engagement in BP monitoring are needed, and future strategies could include a more systematized nudge-based approach in which health professionals regularly provide patients feedback based on their BP measurements. Future work could also explore a combination of remote BP monitoring, health professional feedback, and engaging patients' social support members (ie, family and friends), much like an intervention currently being studied at Penn Family Care [36].

At the health care system level, physicians had several suggestions to address the barriers that surfaced, including leveraging team-based care to sift through patient BP measurements and buoy individualized feedback to patients, AI-based tools to surface clinically relevant data, and improved BP cuff technology. A recent meta-analysis showed that multicomponent strategies, including team-based care and medication titration by a nonphysician (ie, an MA or pharmacist), were most effective for systolic BP reduction compared with other interventions [37]. A study found that most patients are interested in using AI-based tools in their care [38], which may someday include an AI-based algorithm in development that differentiates clinically relevant BP measurements from outliers and extraneous data [39]. In addition, the AHA recently called for improved BP cuff technology and a set of clear standards on how to use these cuffs, echoing this study's issue with technical glitches with the BP cuffs used [40]. Bluetooth-enabled BP cuffs can allow clinicians to monitor patterns in patients' BP data. However, there are several access barriers to consider, including the need for a smartphone app to connect the Bluetooth BP cuff and consistent Wi-Fi, which may not be affordable for all patients or available in rural areas [41]. Future research is needed to explore patients' and health professionals' perceptions of Bluetooth-enabled BP cuffs relative to manual BP cuffs for home monitoring. In addition, implementation science research is necessary to determine whether Bluetooth-enabled BP cuffs can be implemented in various settings given their cost and mixed feasibility.

The COVID-19 pandemic has surfaced a clinical need and demand for remote BP monitoring. According to a study published by the AHA, BP control worsened in both men and women at the onset of the COVID-19 pandemic [42]. A recent qualitative study found that PCPs believe that remote BP monitoring can improve hypertension management, but successful implementation requires improving patient acceptance and seamless integration into clinical workflows [32]. We would expect remote BP monitoring to become the gold standard based on this evaluation and previous research. For remote BP monitoring to be accessible and sustainably used, there would need to be improvements in BP data visualization and EMR incorporation [43], Bluetooth technology [44], insurance

coverage of Bluetooth BP cuff technology [45], and team-based care [46].

Strengths and Limitations

Several limitations should be noted for interpreting the study findings, including testing in a single clinic, the small number of patients and health professionals interviewed, and the absence of quantitative data to assess actual patient changes in BP. Of the 50 patients in the Humanwide pilot, 34 (68%) were not interviewed because of our convenience sampling method. This method has known limitations, including lack of generalization, inability to represent subpopulations accurately, and bias toward people who will participate [47]. To account for these biases, we attended to demographics of the sample, making a purposive attempt to include perspectives from a variety of roles for clinicians and determining that our demographic balance for patients would include all groups of interest.

Thematic saturation has been systematically assessed in previous qualitative studies, with determinations that thematic saturation was reached in 2 studies at 12 interviews [48,49]. As previously mentioned, our sample sizes for health professionals and patients were relatively small (n=15 and n=16). On the basis of previous work with saturation and our assessment of no novel themes regarding BP, we found our patient data set to represent thematic saturation. Our clinician data set confirmed the themes from the patient data set. This agreement between data sets is a strong indicator through data triangulation [50] that our findings represent thematic saturation of BP perspectives in this pilot. All patients who mentioned remote BP monitoring or hypertension (10/16, 63%) were included in the data analysis. This study highlights the experiences of multiple stakeholders, including patients and health professionals, to inform the future dissemination of hypertension management in Humanwide and precision health more broadly.

Conclusions

We found that, of the 4 components of Humanwide (pharmacogenomics, genetic testing, digital health, and health coaching), digital health via remote BP monitoring was reported to be the most impactful for hypertension management from the perspective of both patients and health professionals. Despite the barriers, remote BP monitoring is promising, as reflected by enhanced patient-health professional engagement, perceived improvements in patient care, and increased clinic efficiency. Future recommendations to overcome barriers include the integration of patient BP data into the EMR, automated ways of monitoring and identifying actionable BP measurements, leveraging team-based care to facilitate data monitoring and individualized feedback, and tools to increase sustained patient use.

Acknowledgments

The authors would like to thank all the patients and health professionals for participating in the Humanwide pilot. Humanwide was funded by the Stanford School of Medicine Dean's Office and implemented by Stanford Primary Care in Santa Clara. The Stanford School of Medicine Dean's Office was not a part of the qualitative analysis, including the interpretation of the data, the writing of the manuscript, and the decision to submit the manuscript for publication. No public funding was used for the activities reported in this manuscript. The evaluations reported in this manuscript were financially supported by Stanford Health Care. The manuscript contents are solely the responsibility of the authors and do not necessarily represent the official views of Stanford Health Care, an entity affiliated with but not part of the Stanford School of Medicine where the Evaluation Sciences Unit resides.

Data Availability

Supporting data are available on request. Please contact the corresponding author.

Conflicts of Interest

JB, CBJ, NS, MW, and LGR are affiliated with the Evaluation Sciences Unit. MM is currently a practicing health professional in the Stanford Health Care system.

References

1. Perkovic V, Huxley R, Wu Y, Prabhakaran D, MacMahon S. The burden of blood pressure-related disease. Hypertension 2007 Dec;50(6):991-997. [doi: 10.1161/hypertensionaha.107.095497]
2. Nearly half of U.S. adults could now be classified with high blood pressure, under new definitions. American Heart Association News. URL: https://www.heart.org/en/news/2018/05/01/nearly-half-of-us-adults-could-now-be-classified-with-high-blood-pressure-under-new-definitions [accessed 2022-05-03]
3. High blood pressure fact sheet. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/bloodpressure/facts.htm [accessed 2022-05-03]
4. Law MR, Morris JK, Wald NJ. Use of blood pressure lowering drugs in the prevention of cardiovascular disease: meta-analysis of 147 randomised trials in the context of expectations from prospective epidemiological studies. BMJ 2009 May 19;338:b1665 [FREE Full text] [doi: 10.1136/bmj.b1665] [Medline: 19454737]
5. Underlying cause of death 1999-2020. Centers for Disease Control and Prevention. URL: https://wonder.cdc.gov/wonder/help/ucd.html [accessed 2022-05-03]

6. Savoia C, Volpe M, Grassi G, Borghi C, Agabiti Rosei E, Touyz R. Personalized medicine-a modern approach for the diagnosis and management of hypertension. Clin Sci (Lond) 2017 Nov 15;131(22):2671-2685 [FREE Full text] [doi: 10.1042/CS20160407] [Medline: 29109301]
7. Hajjar I, Kotchen TA. Trends in prevalence, awareness, treatment, and control of hypertension in the United States, 1988-2000. JAMA 2003 Jul 09;290(2):199-206. [doi: 10.1001/jama.290.2.199] [Medline: 12851274]
8. What is precision medicine? MedlinePlus. URL: https://ghr.nlm.nih.gov/primer/precisionmedicine/definition [accessed 2022-05-03]
9. Bíró K, Dombrádi V, Jani A, Boruzs K, Gray M. Creating a common language: defining individualized, personalized and precision prevention in public health. J Public Health (Oxf) 2018 Dec 01;40(4):e552-e559. [doi: 10.1093/pubmed/fdy066] [Medline: 29897560]
10. Mahoney MR, Asch SM. Humanwide: a comprehensive data base for precision health in primary care. Ann Fam Med 2019 May 13;17(3):273 [FREE Full text] [doi: 10.1370/afm.2342] [Medline: 31085532]
11. Tucker KL, Sheppard JP, Stevens R, Bosworth HB, Bove A, Bray EP, et al. Self-monitoring of blood pressure in hypertension: a systematic review and individual patient data meta-analysis. PLoS Med 2017 Sep 19;14(9):e1002389 [FREE Full text] [doi: 10.1371/journal.pmed.1002389] [Medline: 28926573]
12. Margolis KL, Asche SE, Dehmer SP, Bergdall AR, Green BB, Sperl-Hillen JM, et al. Long-term outcomes of the effects of home blood pressure telemonitoring and pharmacist management on blood pressure among adults with uncontrolled hypertension: follow-up of a cluster randomized clinical trial. JAMA Netw Open 2018 Sep 07;1(5):e181617 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.1617] [Medline: 30646139]
13. Parati G, Dolan E, McManus RJ, Omboni S. Home blood pressure telemonitoring in the 21st century. J Clin Hypertens (Greenwich) 2018 Jul 13;20(7):1128-1132 [FREE Full text] [doi: 10.1111/jch.13305] [Medline: 30003701]
14. Fletcher BR, Hinton L, Hartmann-Boyce J, Roberts NW, Bobrovitz N, McManus RJ. Self-monitoring blood pressure in hypertension, patient and provider perspectives: a systematic review and thematic synthesis. Patient Educ Couns 2016 Feb;99(2):210-219. [doi: 10.1016/j.pec.2015.08.026] [Medline: 26341941]
15. Lv N, Xiao L, Simmons ML, Rosas LG, Chan A, Entwistle M. Personalized hypertension management using patient-generated health data integrated with electronic health records (empower-h): six-month pre-post study. J Med Internet Res 2017 Sep 19;19(9):e311 [FREE Full text] [doi: 10.2196/jmir.7831] [Medline: 28928111]
16. Enforcement policy for non-invasive remote monitoring devices used to support patient monitoring during the coronavirus disease 2019 (COVID-19) public health emergency (revised). U.S. Food and Drug Administration. URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/enforcement-policy-non-invasive-remote-monitoring-devices-used-support-patient-monitoring-during [accessed 2022-05-03]
17. HEDIS for hypertension. California Heath & Wellness. URL: https://www.cahealthwellness.com/newsroom/hedis-for-hypertension-.html [accessed 2022-05-03]
18. Unger T, Borghi C, Charchar F, Khan NA, Poulter NR, Prabhakaran D, et al. 2020 International society of hypertension global hypertension practice guidelines. Hypertension 2020 Jun;75(6):1334-1357. [doi: 10.1161/hypertensionaha.120.15026]
19. Whelton PK, Carey RM, Aronow WS, Casey DE, Collins KJ, Dennison Himmelfarb C, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: a report of the American college of cardiology/American heart association task force on clinical practice guidelines. J Am Coll Cardiol 2018 May 15;71(19):e127-e248 [FREE Full text] [doi: 10.1016/j.jacc.2017.11.006] [Medline: 29146535]
20. Palacholla RS, Fischer N, Coleman A, Agboola S, Kirley K, Felsted J, et al. Provider- and patient-related barriers to and facilitators of digital health technology adoption for hypertension management: scoping review. JMIR Cardio 2019 Mar 26;3(1):e11951 [FREE Full text] [doi: 10.2196/11951] [Medline: 31758771]
21. Omboni S, Panzeri E, Campolo L. E-health in hypertension management: an insight into the current and future role of blood pressure telemonitoring. Curr Hypertens Rep 2020 Jun 06;22(6):42. [doi: 10.1007/s11906-020-01056-y] [Medline: 32506273]
22. Brown-Johnson CG, Safaeinili N, Baratta J, Palaniappan L, Mahoney M, Rosas LG, et al. Implementation outcomes of Humanwide: integrated precision health in team-based family practice primary care. BMC Fam Pract 2021 Feb 02;22(1):28 [FREE Full text] [doi: 10.1186/s12875-021-01373-4] [Medline: 33530939]
23. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011 Mar 19;38(2):65-76 [FREE Full text] [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
24. Technology HealthKit. Apple Developer Documentation. URL: https://developer.apple.com/documentation/healthkit [accessed 2022-05-03]
25. Blood pressure monitors. Withings. URL: https://www.withings.com/us/en/blood-pressure-monitors [accessed 2022-05-03]
26. Fereday J, Muir-Cochrane E. Demonstrating rigor using thematic analysis: a hybrid approach of inductive and deductive coding and theme development. Int J Qual Method 2016 Nov 29;5(1):80-92. [doi: 10.1177/160940690600500107]
27. AlYahmady HH, Al Abri SS. Using Nvivo for data analysis in qualitative research. Int Interdisciplinary J Educ 2013;2(2):181-186. [doi: 10.12816/0002914]

28. Donahue KE, Tillman J, Halladay JR, Cené CW, Hinderliter A, Cummings DM, et al. Lessons learned from implementing health coaching in the heart healthy lenoir hypertension study. Prog Community Health Partnersh 2016;10(4):559-567. [doi: 10.1353/cpr.2016.0064] [Medline: 28569681]
29. Postel-Vinay N, Bobrie G, Savard S, Persu A, Amar L, Azizi M, et al. Home blood pressure measurement and digital health: communication technologies create a new context. J Hypertens 2018 Nov;36(11):2125-2131. [doi: 10.1097/HJH.0000000000001860] [Medline: 30063638]
30. Get the most out of home blood pressure monitoring. Mayo Clinic. URL: https://www.mayoclinic.org/diseases-conditions/high-blood-pressure/in-depth/high-blood-pressure/art-20047889 [accessed 2022-05-03]
31. George J, MacDonald T. Home blood pressure monitoring. Eur Cardiol 2015 Dec;10(2):95-101 [FREE Full text] [doi: 10.15420/ecr.2015.10.2.95] [Medline: 30310433]
32. Gondi S, Ellis S, Gupta M, Ellerbeck E, Richter K, Burns J, et al. Physician perceived barriers and facilitators for self-measured blood pressure monitoring- a qualitative study. PLoS One 2021 Aug 20;16(8):e0255578 [FREE Full text] [doi: 10.1371/journal.pone.0255578] [Medline: 34415946]
33. Ruiz-Fernández D, Marcos-Jorquera D, Gilart-Iglesias V, Vives-Boix V, Ramírez-Navarro J. Empowerment of patients with hypertension through BPM, IoT and remote sensing. Sensors (Basel) 2017 Oct 04;17(10):2273 [FREE Full text] [doi: 10.3390/s17102273] [Medline: 28976940]
34. Etminani K, Tao Engström A, Göransson C, Sant'Anna A, Nowaczyk S. How behavior change strategies are used to design digital interventions to improve medication adherence and blood pressure among patients with hypertension: systematic review. J Med Internet Res 2020 Apr 09;22(4):e17201 [FREE Full text] [doi: 10.2196/17201] [Medline: 32271148]
35. Omboni S. Connected health in hypertension management. Front Cardiovasc Med 2019 Jun 13;6:76 [FREE Full text] [doi: 10.3389/fcvm.2019.00076] [Medline: 31263703]
36. Remote monitoring and social support for hypertension management (SupportBP). ClinicalTrials. URL: https://clinicaltrials.gov/ct2/show/NCT03416283 [accessed 2022-05-03]
37. Mills KT, Obst KM, Shen W, Molina S, Zhang H, He H, et al. Comparative effectiveness of implementation strategies for blood pressure control in hypertensive patients. Ann Intern Med 2017 Dec 26;168(2):110. [doi: 10.7326/m17-1805]
38. Tran V, Riveros C, Ravaud P. Patients' views of wearable devices and AI in healthcare: findings from the ComPaRe e-cohort. NPJ Digit Med 2019 Jun 14;2(1):53 [FREE Full text] [doi: 10.1038/s41746-019-0132-y] [Medline: 31304399]
39. The health care benefits of combining wearables and AI. Harvard Business Review. URL: https://hbr.org/2019/05/the-health-care-benefits-of-combining-wearables-and-ai [accessed 2022-05-03]
40. Muntner P, Shimbo D, Carey RM, Charleston JB, Gaillard T, Misra S, et al. Measurement of blood pressure in humans: a scientific statement from the American heart association. Hypertension 2019 May;73(5):e35-e66 [FREE Full text] [doi: 10.1161/HYP.0000000000000087] [Medline: 30827125]
41. Goldberg EM, Levy PD. New approaches to evaluating and monitoring blood pressure. Curr Hypertens Rep 2016 Jun;18(6):49 [FREE Full text] [doi: 10.1007/s11906-016-0650-9] [Medline: 27137524]
42. Laffin LJ, Kaufman HW, Chen Z, Niles JK, Arellano AR, Bare LA, et al. Rise in blood pressure observed among US adults during the COVID-19 pandemic. Circulation 2022 Jan 18;145(3):235-237. [doi: 10.1161/circulationaha.121.057075]
43. McGreevey JD, Mallozzi CP, Perkins RM, Shelov E, Schreiber R. Reducing alert burden in electronic health records: state of the art recommendations from four health systems. Appl Clin Inform 2020 Jan 01;11(1):1-12 [FREE Full text] [doi: 10.1055/s-0039-3402715] [Medline: 31893559]
44. Piper MA, Evans CV, Burda BU, Margolis KL, O'Connor E, Whitlock EP. Diagnostic and predictive accuracy of blood pressure screening methods with consideration of rescreening intervals: a systematic review for the U.S. preventive services task force. Ann Intern Med 2015 Feb 03;162(3):192-204. [doi: 10.7326/m14-1539]
45. CMS expands coverage of Ambulatory Blood Pressure Monitoring (ABPM). Centers for Medicare & Medicaid Services. URL: https://www.cms.gov/newsroom/press-releases/cms-expands-coverage-ambulatory-blood-pressure-monitoring-abpm [accessed 2022-05-03]
46. Lee NS, Anastos-Wallen R, Chaiyachati KH, Reitz C, Asch DA, Mehta SJ. Clinician decisions after notification of elevated blood pressure measurements from patients in a remote monitoring program. JAMA Netw Open 2022 Jan 04;5(1):e2143590 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.43590] [Medline: 35029664]
47. Andrade C. The inconvenient truth about convenience and purposive samples. Indian J Psychol Med 2021 Jan 17;43(1):86-88 [FREE Full text] [doi: 10.1177/0253717620977000] [Medline: 34349313]
48. Ando H, Cousins R, Young C. Achieving saturation in thematic analysis: development and refinement of a codebook. Comprehensive Psychol 2014 Feb 20;3:03.CP.3.4. [doi: 10.2466/03.cp.3.4]
49. Guest G, Bunce A, Johnson L. How many interviews are enough? Field Methods 2016 Jul 21;18(1):59-82. [doi: 10.1177/1525822X05279903]
50. Johnson M, O'Hara R, Hirst E, Weyman A, Turner J, Mason S, et al. Multiple triangulation and collaborative research using qualitative methods to explore decision making in pre-hospital emergency care. BMC Med Res Methodol 2017 Jan 24;17(1):11 [FREE Full text] [doi: 10.1186/s12874-017-0290-z] [Medline: 28118817]

Abbreviations

AHA: American Heart Association
AI: artificial intelligence
BP: blood pressure
EMR: electronic medical record
MA: medical assistant
MD: medical doctor
PCP: primary care physician

Edited by A Mavragani; submitted 18.08.21; peer-reviewed by B Green, J Edwards, L Weinert; comments to author 11.01.22; revised version received 07.03.22; accepted 13.04.22; published 10.06.22.

Please cite as:
Baratta J, Brown-Johnson C, Safaeinili N, Goldman Rosas L, Palaniappan L, Winget M, Mahoney M
Patient and Health Professional Perceptions of Telemonitoring for Hypertension Management: Qualitative Study
JMIR Form Res 2022;6(6):e32874
URL: https://formative.jmir.org/2022/6/e32874
doi:10.2196/32874
PMID:35687380

Original Paper

Online Partner Seeking and Sexual Behaviors Among Men Who Have Sex With Men From Small and Midsized Towns: Cross-sectional Study

Vira Pravosud[1], PhD; April M Ballard[2], MPH; Ian W Holloway[3], PhD; April M Young[4,5], PhD

[1]Center for Tobacco Control Research and Education, Cardiovascular Research Institute, University of California, San Francisco, San Francisco, CA, United States

[2]Gangarosa Department of Environmental Health, Emory University Rollins School of Public Health, Atlanta, GA, United States

[3]Luskin School of Public Affairs, University of California, Los Angeles, Los Angeles, CA, United States

[4]Department of Epidemiology, College of Public Health, University of Kentucky, Lexington, KY, United States

[5]Center on Drug and Alcohol Research, Department of Behavioral Science, University of Kentucky College of Medicine, Lexington, KY, United States

Corresponding Author:
Vira Pravosud, PhD
Center for Tobacco Control Research and Education
Cardiovascular Research Institute
University of California, San Francisco
530 Parnassus Avenue, Suite 366
San Francisco, CA, 94143
United States
Phone: 1 415 514 8627
Email: vira.pravosud@ucsf.edu

Abstract

Background: Men who have sex with men (MSM) residing outside of large urban areas are underrepresented in research on online partner seeking and sexual behaviors related to transmission of HIV.

Objective: We aimed to determine associations between the use of the internet or social networking apps (*online tools*) to meet partners for sex, dating, or for both purposes (*online partner seeking*) and sexual behaviors among MSM residing in small and midsized towns in Kentucky, United States.

Methods: Using peer-referral sampling and online self-administered questionnaires, data were collected from 252 men, aged 18 to 34 years, who had recently (past 6 months) engaged in anal sex with another man and resided in Central Kentucky. Using multivariable logistic regression models, we assessed associations of online partner seeking and HIV-related sexual behaviors.

Results: Most (181/252, 71.8%) of the participants reported using online tools for partner seeking. Of these 181 respondents, 166 (91.7%) had used online tools to meet partners for sex (n=45, 27.1% for sex only; and n=121, 72.9% for sex and dating) and 136 (75.1%) had used online tools to meet partners for dating (n=15, 11% for dating only; and n=121, 89% for sex and dating). Adjusted analyses revealed that MSM who had engaged in condomless insertive and receptive anal intercourse were less likely to report online partner seeking (adjusted odds ratio [aOR] 0.22, 95% CI 0.07-0.68; *P*=.009 and aOR 0.25, 95% CI 0.10-0.66; *P*=.005, respectively). Increased number of insertive and receptive anal sex partners and substance use before or during sex were associated with higher odds of online partner seeking (aOR 1.31, 95% CI 1.11-1.55; *P*=.001; aOR 1.20, 95% CI 1.05-1.39; *P*=.008; and aOR 2.50, 95% CI 1.41-4.44; *P*=.002, respectively).

Conclusions: Among MSM who reside outside of large urban areas and practice online partner seeking, HIV risk-reduction interventions should address safer sex practices, including the risks for HIV transmission associated with alcohol or drug use before or during sex. MSM who do not practice online partner seeking are in need of continued outreach to reduce condomless anal sex.

(JMIR Form Res 2022;6(6):e35056) doi:10.2196/35056

KEYWORDS

men who have sex with men; MSM; sexual risk behaviors; social networking and dating apps; online tools; HIV; sexually transmitted infection; STI prevention; mobile phone

Introduction

Background

Men who have sex with men (MSM) have the highest burden of HIV in the United States [1], with almost 70% of the new HIV cases being attributable to male-to-male sexual contact [1]. Although prior research has contributed immensely to our understanding of prevention of HIV transmission among MSM, new avenues for partner seeking among MSM have emerged over the past 3 decades, including through the internet, social media, and geosocial networking apps.

The use of the internet and social media can facilitate health-protective and safer sex practices among MSM who seek partners online [2,3]. Internet profiles become a place for MSM to disclose attitudes toward substance use [4] and HIV status [2,4-7], as well as inform others whether they take pre-exposure prophylaxis (PrEP) for HIV prevention [8]. Such practices may help internet-using MSM to make more informed sexual health decisions [2,5,7]. As described in previous research some internet-using MSM are more likely to report HIV testing [9] and condom use during anal intercourse with partners met online [10-12].

However, online partner seeking has also been associated with increased engagement in HIV-related risk behaviors [9-11,13-20], including an increased number of sex partners [11,17,18,21], higher prevalence of substance use [9,22], condomless anal intercourse with male partners [13,18,19], and casual sex with HIV-positive partners met online [16]. Compared with meeting partners in person, online partner seeking among MSM has also been associated with potential transmission of HIV [23] and greater odds of testing positive for sexually transmitted infections (STIs; eg, syphilis [24], gonorrhea, and chlamydia [22]), but the factors driving these associations are not well understood.

Most studies to date have described associations between the use of online partner seeking and sexual risk behaviors among MSM residing in large metropolitan centers; for example, Los Angeles [11,22,25], New York City [7,18,26,27], and San Francisco [28] in the United States. Few studies (including those published more than 10 years ago [29-31]) have examined patterns of internet use and associated sexual risk behaviors among MSM from rural areas [29-32] or small and midsized towns [33]. Similar to findings from large metropolitan areas [22], rural MSM who used the internet to search for sex partners had a higher prevalence of condomless anal intercourse [30,31]. Evidence shows that MSM from rural communities can be especially susceptible to social estrangement [34] and hostility [35], as well as sexual isolation and stigma [36,37], and that rural MSM often have few identifiable venues where they can meet other MSM [33,35,38,39]. MSM who reside outside of large urban areas may use social networking and dating websites or mobile apps as a safe and convenient way to meet partners [35] and travel from, or to, nearby metropolitan centers to meet people contacted online [35,40,41]. These findings highlight the need for additional research on sexual behaviors and the use of various online tools for partner seeking among MSM across the urban and nonurban continuum.

The Goal of This Study

As noted [42], MSM from small and midsized towns have been underrepresented in research, leaving a gap in knowledge that is concerning especially among those who reside in the American South, a region that remains disproportionately affected by HIV [1,43]. As there is a need for more geographic diversity in studies on the topic, we focused on MSM residing in 15 Central Kentucky counties that predominantly consist of small and midsized towns [44,45] and are located in the Bluegrass Area Development District [46]. In 2019, the Bluegrass Area Development District had the second greatest percentage of HIV diagnoses (19%) in the state of Kentucky [47], where the majority of cases among men were diagnosed in MSM (67%), persons who inject drugs (PWID; 8%), or MSM and PWID (6%) [47]. The relevance of the region to HIV research also lies in the region's geographic proximity to rural Central Appalachia, which is considered highly vulnerable to an HIV outbreak associated with injection drug use [48], and to 2 recent HIV clusters in Northern Kentucky [49,50] and West Virginia [51], wherein most cases were among PWID [52] and PWID with male-to-male sexual contact [53]. In 2020, Kentucky was named 1 of just 7 rural states targeted by the Ending the HIV Epidemic initiative based on the percentage of new diagnoses occurring in rural areas [54]. Therefore, the aim of this exploratory study was to describe online partner seeking among MSM residing in small and midsized towns in Kentucky and to compare HIV sexual risk behaviors in this understudied population between those who use online tools for partner seeking and those who do not.

Methods

Study Design and Data Collection

From February 2018 to July 2018, a total of 253 participants were recruited using targeted street outreach and web-based respondent-driven sampling, which was previously shown to be an effective technique for reaching stigmatized populations, including MSM [55]. The recruitment process for the study is described in more detail elsewhere [56]. Briefly, the research team posted flyers on social media through a study-specific page and young adult lesbian, gay, bisexual, transgender, and queer groups and distributed flyers at various local venues that were outwardly friendly to members of the lesbian, gay, bisexual, transgender, and queer community (including adult entertainment stores, bars, and health clinics) as well as at 2 local pride festivals. All flyers directed individuals to the study website, where a link to the study's online Qualtrics survey was posted. Upon completion of the survey, participants were offered a US $25 incentive in the form of an e-gift card or mailed payment [56]. Participants could refer up to 3 peers and receive

up to US $30 if their referrals were eligible and completed the survey [56].

On the basis of demographic and behavioral factors associated with new HIV infections, the eligibility criteria included being male, aged 18 to 34 years, having engaged in anal sex with another man during the past 6 months, and residing in 1 of 15 Central Kentucky counties. The study specifically focused on young MSM because they experience the highest burden of HIV [57]. In total, 787 screening questionnaires were received, among which 490 (62.2%) were eligible and 414 (52.6%) were complete survey entries. Recent studies have demonstrated that online research may be at risk of fraudulent or invalid responses [58,59]. As such, the study staff implemented a rigorous fraud prevention and detection mechanism to ensure data quality based on check of geolocation, telephone number, name, email address, personal information, and survey duration [56]. The protocol also included direct outreach by staff to those deemed potentially fraudulent. Of the 414 responses considered, after implementing our fraud-detection strategy, we excluded 161 (38.9%) responses, leaving 253 (61.1%) valid responses [56]. The 253 participants screened eligible by checking that they had engaged in anal sex with another man in the past 6 months. However, during further data cleaning, we identified that, of these 253 participants, 1 (0.4%) contradicted their screening answer in their response of a zero value for the question about the number of lifetime male partners. Thus, this participant was excluded. The final analytic sample included 252 participants who were recruited online, through flyers, or through in-person outreach activities.

Ethics Approval

The institutional review board at the University of Kentucky approved the study protocol (76147). Survey participants completed informed consent forms [56].

Measures

Outcome Variable

As many social networking and dating sites have both webpages and smartphone-based apps, online partner seeking in this study is defined as the use of either type of online platform for dating or sex. The outcome was defined as an affirmative response to at least one of the following 2 survey questions: “In the past 6 months, have you used any social media tools or apps to meet people for sex?” and “In the past 6 months, have you used any social media tools or apps to meet partners to date, not necessarily just for sex?” Given their significant correlation (χ^2_1=70.1; P<.001), the answers were combined in an aggregate outcome: online partner seeking, which was a binary variable (yes or no).

Demographic Characteristics

Participants provided information on various demographic characteristics, including age, gender, race, ethnicity, sexual orientation, highest level of education completed, income in the past 30 days (in US $), and county of residence. Because of the sample’s limited heterogeneity on gender, race and ethnicity, sexual orientation, and education, these variables were recoded into 2-level categories representing, respectively, man versus other gender groups; White and non-Hispanic versus other racial and ethnic groups; gay or mostly gay versus other sexual orientation; high school or General Educational Development Test (GED) or less versus (some) college degree or higher. Rural-Urban Continuum Codes [44] were used to classify counties of residence as metropolitan (5 counties, codes 1-3) and nonmetropolitan (6 counties, codes 4-6). Of note, the metropolitan counties were far smaller than the metropolises that have been the focus of previous research (eg, Los Angeles and San Francisco); the largest county had a population of less than 325,000 [45] and encompassed a midsized town. The other 4 metropolitan counties ranged in population size from 24,939 to 48,586 [44].

Behavioral Characteristics and STI and HIV Status

The survey included questions about lifetime behaviors such as sex with HIV-positive partners, HIV testing, antiretroviral therapy, and PrEP use among the respondents. The survey also asked about recent (past 6 months) drug use and engagement in sexual behaviors, including types of intercourse (insertive only, receptive only, and either insertive or receptive), number of male anal sex partners with whom the participant was the insertive (ie, was top) or receptive (ie, was bottom) partner, and frequency of condom use in receptive and insertive positions. Condomless insertive and receptive anal intercourse were analyzed as 2 binary variables created based on responses to questions that asked participants to report the percentage of recent anal sex acts in which a condom was used. Because of kurtosis in the distribution for continuous variables as well as the HIV risk entailed in any condomless sex [60], the variables were dichotomized where condomless insertive or receptive anal intercourse was defined as <100% condom use during insertive or receptive sex acts in the past 6 months (similar to previous research [61,62]). Of note, the survey did not ask participants to specify the HIV status of their partners. In addition, we examined recent engagement in group sex and use of alcohol or any drug (including marijuana, prescription drugs to get high, or any other illicit drug) before or during sex. The survey also asked whether participants had ever been diagnosed with STIs (chlamydia, gonorrhea, oral or anal herpes, or syphilis) or tested positive for HIV in their lifetime.

Statistical Analysis

SAS software (version 9.4; SAS Institute Inc) was used to conduct all analyses. Unadjusted odds ratios (ORs) and adjusted ORs (aORs) and 95% CIs obtained from bivariate and multivariable logistic regression models, respectively, were used as the measures of association between online partner seeking and demographic and behavioral characteristics. We estimated 3 models examining the association between online partner seeking and 3 behaviors relevant to HIV transmission based on previous literature [60,63-66] that were significantly (P<.05) associated with online partner seeking in bivariate analyses and had a sufficient cell size for analysis: condomless insertive anal intercourse (model 1) and condomless receptive anal intercourse (model 2) with male partners, as well as substance use before or during sex (model 3).

On the basis of findings from previous research about differences in online partner seeking by demographics as well

as results of diagnostics of confounding by demographic variables in our analysis, final models were adjusted for age [9,67], education [9], and race and ethnicity [9,15,25,68,69]. Models focused on estimating the association between online partner seeking and condomless insertive or receptive anal intercourse were also adjusted for the number of male insertive or receptive partners in the past 6 months and PrEP use in the lifetime. We examined all possible 2-way interactions for each model with their corresponding covariates and found no significant interactions. In addition, we tested for multicollinearity in all adjusted models and identified no multicollinearity. We also conducted sensitivity analyses while adjusting all models for residence in metropolitan and nonmetropolitan counties.

We used complete case analysis for each of the final multivariable regression models. Examination of missing versus included observations revealed no associations between missingness and most of the covariates of interest and no differences regarding the values of the dependent variable, online partner seeking.

Results

Characteristics of Survey Participants

Of the 252 respondents, 205 (81.3%) were White and non-Hispanic, 244 (96.8%) identified as male, and 215 (85.3%) self-identified as gay or mostly gay. The median age of participants was 25.8 (IQR 22.8-29.2) years. Of the 252 respondents, 216 (85.7%) were in college or had obtained an undergraduate degree or higher; in addition, 214 (84.9%) were currently residing in a metropolitan county, whereas 38 (15.1%) resided in nonmetropolitan areas (Table 1).

Of the 181 respondents who reported using online tools for partner seeking, 166 (91.7%) had used online tools to meet partners for sex (n=45, 27.1% for sex only; and n=121, 72.9% for sex and dating) and 136 (75.1%) had used online tools to meet partners for dating (n=15, 11% for dating only; and n=121, 89% for sex and dating). Overall, 71.8% (181/252) of the respondents had used online tools to meet other men for sex or dating: 24.9% (45/181) used online tools to meet people for sex only, 8.3% (15/181) used online tools to meet people exclusively for dating, and 66.9% (121/181) used online tools for both sex and dating. More than 20 online tools were listed by the 181 MSM who practiced online partner seeking, including Grindr (n=154, 85.1%), Tinder (n=93, 51.4%), Facebook (n=57, 31.5%), Scruff (n=53, 29.3%), and Adam4Adam (n=28, 15.5%), as well as others such as Backpage, Bumble, Craigslist, eHarmony, GROWLr, Hornet, Hot or Not, Instagram, Jack’d, Manhunt, Match, Meetme, OK Cupid, Plenty of Fish, Snapchat, Tumblr, Twitter, and VGL.

Table 1. Demographic, health, and behavioral characteristics of the respondents (N=252).

Characteristics or behavior	Values[a]
Demographic characteristics	
Age (years), median (IQR)	25.8 (22.8-29.2)
Income in the past 30 days (US $), median (IQR)	1600 (800-2500)
Gender, n (%)	
Man	244 (96.8)
Other (transgender man or woman; gender queer)	8 (3.2)
Race, n (%)	
White	211 (83.7)
Black or African American	20 (7.9)
Other (Pacific Islander, biracial, or multiracial)	18 (7.1)
Ethnicity, n (%)	
Non-Hispanic	235 (93.3)
Hispanic	12 (4.8)
Race and ethnicity, n (%)	
White and non-Hispanic	205 (81.3)
Other	40 (15.9)
Sexual orientation, n (%)	
Gay or mostly gay	215 (85.3)
Other (bisexual or heterosexual or pansexual or queer)	36 (14.3)
Education, n (%)	
Less than or some high school or high school graduate or GED[b]	36 (14.3)
Some college or college graduate or higher	216 (85.7)
County of residence[c], n (%)	
Metropolitan	214 (84.9)
Nonmetropolitan	38 (15.1)
Recent behaviors (past 6 months)	
Type of anal sex with male partners, n (%)	
Only insertive (ie, participant was always in insertive position)	59 (23.4)
Only receptive (ie, participant was always in receptive position)	44 (17.5)
Insertive or receptive	148 (58.7)
Number of male insertive anal sex partners[d], median (IQR)	1 (1-3)
Number of male receptive anal sex partners[e], median (IQR)	1 (1-4)
Number of male oral sex partners, median (IQR)	3 (1-8)
Condomless insertive anal intercourse[d], n (%)	180 (71.4)
Condomless receptive anal intercourse[e], n (%)	170 (67.5)
Engagement in group sex, n (%)	63 (25)
Recent substance use (past 6 months), n (%)	
Used drugs to get high	144 (57.1)
Daily use of drugs to get high	33 (13.1)
Alcohol or any illicit drug use before or during sex	151 (59.9)

Characteristics or behavior	Values[a]
Illicit drug use before or during sex with people met through social media (N=181; not available: 71)	65 (35.9)
Lifetime behaviors, n (%)	
Ever had sex with HIV-positive partners	54 (21.4)
Ever used drugs to get high in the lifetime	192 (76.2)
Ever injected drugs in the lifetime	14 (5.6)
HIV or STI[f] status and PrEP[g] use, n (%)	
Ever tested positive for STIs	72 (28.6)
Ever been tested for HIV	215 (85.3)
Ever tested positive for HIV	17 (6.7)
Antiretroviral therapy uptake	17 (6.7)
Ever used PrEP	37 (14.7)[h]

[a]Percentages are out of the total sample and may not add to total 100% because of rounding to 1 decimal. Sample sizes for some variables could vary because of missing data as well as refuse to answer or not applicable (n/a) responses.

[b]GED: General Educational Development Test.

[c]The US Department of Agriculture 2013 Rural-Urban Continuum Codes were used to classify counties of residence as metropolitan (5 counties, codes 1-3) and nonmetropolitan (6 counties, codes 4-6).

[d]With whom participants were in the insertive position.

[e]With whom participants were in the receptive position.

[f]STI: sexually transmitted infection.

[g]PrEP: pre-exposure prophylaxis.

[h]PrEP use in the lifetime was reported by 1 respondent with self-reported HIV-positive status.

Unadjusted Associations of Online Partner Seeking

Some health-protective behaviors were more common among those using online tools (Table 2). The odds of reporting online partner seeking were lower for those who practiced condomless insertive (OR 0.26, 95% CI 0.10-0.70; *P*=.008) and receptive (OR 0.30, 95% CI 0.12-0.75; *P*=.01) anal intercourse with male partners. The odds of reporting online partner seeking were significantly higher among those who reported PrEP use in the lifetime (OR 3.73, 95% CI 1.27-10.96; *P*=.02).

Compared with nonusers, users of online tools for partner seeking reported a higher median number of recent male insertive anal sex (2 vs 1; OR 1.28, 95% CI 1.10-1.49; *P*=.001), receptive anal sex (2 vs 1; OR 1.19, 95% CI 1.05-1.33; *P*=.005), and oral sex partners (4 vs 1; OR 1.25, 95% CI 1.12-1.40; *P*<.001). In our sample, of the 252 respondents, 144 (57.1%) reported drug use in the past 6 months (Table 1); of these 144 respondents, 33 (22.9%) reported daily drug use. Online partner seeking was more likely to be observed among those who reported recent drug use (OR 2.04, 95% CI 1.16-3.57; *P*=.01) as well as substance use before or during sex (OR 2.34, 95% CI 1.34-4.09; *P*=.003). Of the 181 users of online tools, 65 (35.9%) reported recent substance use before or during sex with people they met through social media (Table 1). Although not considered for further analysis because of small cell sizes, engagement in group sex was significantly higher among users than among nonusers of online tools for partner seeking (OR 2.93, 95% CI 1.36-6.32; *P*=.006). Differences in reporting history of HIV (OR 0.87, 95% CI 0.29-2.58; *P*=.80) or STI (OR 1.51, 95% CI 0.79-2.87; *P*=.21) diagnoses were insignificant between users and nonusers.

Table 2. Unadjusted and adjusted comparisons of men who have sex with men who are users versus nonusers of online tools for partner seeking by demographic and behavioral characteristics (N=252).

Characteristics or behavior	Users[a,b] (n=181)	Nonusers[a] (n=71)	OR[c] (95% CI)	*P* value
Demographic characteristics				
Age (years), median (IQR)	25.2 (22.5-28.8)	27.1 (24.1-29.5)	0.94 (0.88-1.00)	.07
Income in the past 30 days (US $), median (IQR)	1500 (750-2500)	1750 (1000-2500)	0.99 (0.98-1.00)[d]	.04
Gender, n (%)				
Man	176 (97.2)	68 (95.8)	Reference	N/A[e]
Other	5 (2.8)	3 (4.2)	0.64 (0.15-2.77)	.55
Race and ethnicity, n (%)				
White and non-Hispanic	145 (82.4)	60 (87)	Reference	N/A
Other	31 (17.6)	9 (13)	1.43 (0.64-3.17)	.39
Sexual orientation, n (%)				
Gay or mostly gay	152 (84)	63 (90)	Reference	N/A
Other (bisexual or heterosexual or pansexual or queer)	29 (16)	7 (10)	1.71 (0.72-4.12)	.23
Education, n (%)				
≤GED[f] or high school graduate	25 (13.8)	11 (15.5)	0.87 (0.41-1.89)	.73
≥Some college or college graduate	156 (86.2)	60 (84.5)	Reference	N/A
County of residence[g], n (%)				
Metropolitan	159 (87.9)	55 (77.5)	2.10 (1.03-4.29)	.04
Nonmetropolitan	22 (12.2)	16 (22.5)	Reference	N/A
Recent behaviors (past 6 months)				
Type of anal sex with male partners, n (%)				
Only insertive (ie, participant was always in insertive position)	44 (24.4)	15 (21.1)	Reference	N/A
Only receptive (ie, participant was always in receptive position)	25 (13.9)	19 (26.8)	0.45 (0.19-1.04)	.06
Insertive or receptive	111 (61.7)	37 (52.1)	1.02 (0.51-2.05)	.95
Number of male insertive anal sex partners[h], median (IQR)	2 (1-5)	1 (0-1)	1.28 (1.10-1.49)	.001
Number of male receptive anal sex partners[i], median (IQR)	2 (1-5)	1 (1-1)	1.19 (1.05-1.33)	.005
Number of male oral sex partners, median (IQR)	4 (2-10)	1 (1-2)	1.25 (1.12-1.40)	<.001
Condomless insertive anal intercourse[h], n (%)	122 (75.3)	58 (92.1)	0.26 (0.10-0.70)	.008
Condomless receptive anal intercourse[i], n (%)	115 (73.3)	55 (90.2)	0.30 (0.12-0.75)	.01
Engagement in group sex, n (%)	54 (29.8)	9 (12.7)	2.93 (1.36-6.32)	.006
Recent substance use (past 6 months) , n (%)				
Used drugs to get high	113 (62.4)	31 (44.9)	2.04 (1.16-3.57)	.01
Alcohol or any illicit drug use before or during sex	119 (65.8)	32 (45.1)	2.34 (1.34-4.09)	.003
Lifetime behaviors, n (%)				
Ever tested positive for STIs[j]	56 (31.3)	16 (23.2)	1.51 (0.79-2.87)	.21
Have been tested for HIV	158 (87.3)	57 (80.3)	1.69 (0.81-3.50)	.16
Ever tested positive for HIV	12 (7.7)	5 (8.8)	0.87 (0.29-2.58)	.80
Antiretroviral therapy uptake	12 (6.7)	5 (7)	0.95 (0.32-2.80)	.92
PrEP[k] use	33 (18.2)	4 (5.6)	3.73 (1.27-10.96)	.02

[a]Percentages may not add to total 100% because of rounding to 1 decimal. Sample sizes for some variables could vary because of missing data as well as refuse to answer or not applicable (n/a) responses.

[b]Includes all respondents who answered *yes* to the questions about using online tools to meet partners for sex, dating, or for both purposes (n=181).

[c]OR: odds ratio.

[d]Per US $100 increase.

[e]N/A: not applicable.

[f]GED: General Educational Development Test.

[g]US Department of Agriculture 2013 Rural-Urban Continuum Codes were used to classify counties of residence as metropolitan (5 counties, codes 1-3) and nonmetropolitan (6 counties, codes 4-6).

[h]With whom participants were in the insertive position.

[i]With whom participants were in the receptive position.

[j]STI: sexually transmitted infection.

[k]PrEP: pre-exposure prophylaxis.

Adjusted Associations of Online Partner Seeking

Adjusting for demographics (Table 3), PrEP use in the lifetime as well as the number of recent male insertive or receptive anal sex partners and condomless insertive and receptive anal intercourse events were associated with lower odds of online partner seeking (aOR 0.22, 95% CI 0.07-0.68; *P*=.009 and aOR 0.25, 95% CI 0.10-0.66; *P*=.005, respectively). Adjusting for demographics, substance use before or during sex was associated with increased odds for online partner seeking (aOR 2.50, 95% CI 1.41-4.44; *P*=.002).

Our sensitivity analysis, while adjusting for residence in metropolitan and nonmetropolitan counties, provided similar results regarding significance and magnitude of associations, indicating lower odds of online partner seeking among those who had reported condomless insertive or receptive anal sex and higher odds of online partner seeking for those who reported substance use before or during sex (Multimedia Appendix 1). Sensitivity analyses also showed that there were no differences in online partner seeking between residents of metropolitan counties and those of nonmetropolitan counties.

Table 3. Adjusted comparisons of men who have sex with men who are users versus nonusers of online tools for partner seeking by sexual and drug-related behavioral characteristics.

Characteristic or behavior	Model 1 (n=214)		Model 2 (n=209)		Model 3 (n=245)	
	aOR[a] (95% CI)	*P* value	aOR (95% CI)	*P* value	aOR (95% CI)	*P* value
Demographic characteristics and PrEP[b] use						
Age	0.89 (0.82-0.97)	.007	0.96 (0.89-1.04)	.34	0.94 (0.87-1.00)	.06
Race and ethnicity						
White and non-Hispanic	Reference	N/A[c]	Reference	N/A	Reference	N/A
Other	0.77 (0.29-2.06)	.61	0.91 (0.35-2.38)	.85	1.40 (0.61-3.23)	.43
Education						
≤GED[d] or high school graduate	0.46 (0.16-1.35)	.16	1.10 (0.39-3.10)	.86	0.64 (0.27-1.56)	.33
≥Some college or college graduate	Reference	N/A	Reference	N/A	Reference	N/A
Ever used PrEP in the lifetime						
Yes	4.26 (1.32-13.77)	.02	3.67 (1.00-13.51)	.05	N/A	N/A
No	Reference	N/A	Reference	N/A	N/A	N/A
Recent behaviors (past 6 months)						
Number of male insertive anal sex partners[e]	1.31 (1.11-1.55)	.001	N/A	N/A	N/A	N/A
Number of male receptive anal sex partners[f]	N/A	N/A	1.20 (1.05-1.39)	.008	N/A	N/A
Condomless insertive anal intercourse[e]	0.22 (0.07-0.68)	.009	N/A	N/A	N/A	N/A
Condomless receptive anal intercourse[f]	N/A	N/A	0.25 (0.10-0.66)	.005	N/A	N/A
Alcohol or any illicit drug use before or during sex	N/A	N/A	N/A	N/A	2.50 (1.41-4.44)	.002

[a]aOR: adjusted odds ratio.

[b]PrEP: pre-exposure prophylaxis.

[c]N/A: not applicable.

[d]GED: General Educational Development Test.

[e]With whom participants were in the insertive position.

[f]With whom participants were in the receptive position.

Discussion

Principal Findings

This cross-sectional study of 252 young adult MSM residing in small and midsized towns in Central Kentucky revealed that 181 (71.8%) respondents had used online tools to meet partners for sex, dating, or for both purposes. Of these 181 respondents, 166 (91.7%) had used online tools to meet partners for sex (n=45, 27.1%, for sex only and n=121, 72.9%, for sex and dating) and 136 (75.1%) had used online tools to meet partners for dating (n=15, 11%, for dating only and n=121, 89%, for sex and dating). In our study, those who practiced online partner seeking reported a higher number of male insertive and receptive anal sex partners in the past 6 months and many reported substance use before or during sex. MSM who used online tools for partner seeking were more likely to report condom use during anal intercourse with male partners than their counterparts who did not practice online partner seeking.

The percentage of men using online tools for partner seeking in our study (181/252, 71.8%) is at the high end of what has been observed in other studies in large urban settings (64.6%-72.1%) [10,20,70]. These findings are consistent with previous research suggesting that because of the limited number of MSM-friendly venues for entertainment and social interaction [33,39], MSM residing outside of large metropolitan areas might use online tools for partner seeking even more to lower the risk of stigma [33,35,36,38] associated with their sexual identity. It is worth noting that our study sample consisted of a substantial proportion of MSM residing outside of major urban areas compared with samples from other studies on the topic that have almost exclusively focused on large urban settings [7,11,18,22,25-28]. Interestingly, despite higher proportions of users of online tools among MSM residing in metropolitan counties, our adjusted analysis revealed no differences in online partner seeking between residents of metropolitan areas and those of nonmetropolitan areas.

In line with previous research [20], including among rural MSM [71], online partner seeking for dating was less common than that for sex among users of online tools in our study: 24.9% (45/181) reported use of online tools exclusively for sex and only 8.3% (15/181) used online tools to seek partners specifically for dating (ie, not necessarily just for sex). Most of the users of online tools (121/181, 66.9%) were interested in seeking partners for both dating and sex. It is likely that MSM who reported online partner seeking for both purposes were predominantly interested in casual or one-time sex partners and, as seen in other studies, were also "dating for fun" [72]. Previous research shows that engagement in risk behaviors may vary by the intention of online partner seeking [73]. Bauermeister et al [73] showed that MSM who frequently practiced online partner seeking specifically for casual sex engaged more in condomless sex acts than MSM who frequently sought dating partners or MSM who rarely practiced online partner seeking. Some authors [14,18,26] have presumed that people who use online tools to look for sexual encounters, especially for casual sex partners, may have higher levels of sensation seeking [74] associated with lower levels of self-control [75]: the 2 characteristics that are generally related to sexual risk taking among MSM [66,75-80]. Because of the small numbers of MSM who reported use of online tools exclusively for dating or for sex, there was insufficient statistical power to detect differences in associations of risk behaviors with online partner seeking by the 2 intentions and future studies are recommended.

Consistent with previous research from large urban centers [10,11,22,67], online partner seeking among MSM in our study was associated with several risk behaviors, including an increased number of male insertive and receptive anal sex partners in the past 6 months before and after adjusting for PrEP use in the lifetime as well as substance use before or during sex. Similar to previous studies [81,82], unadjusted analyses revealed that a larger proportion of users of online tools reported recent drug use and engagement in group sex. Although online partner seeking was associated with various risk factors, our study also showed that some protective behaviors were more prevalent among users of online tools versus nonusers. This finding is important given that some HIV care providers have negative perceptions of MSM who report online partner seeking, and who are assumed to engage in increased sexual risk behaviors, without acknowledgment that it may also be associated with protective behaviors [83]. This study's findings and those of others [9,67,81,82,84,85] demonstrating the association of online partner seeking not only with sexual risk but also with protective behaviors is important to help HIV care providers understand the nuances of online partner seeking and thereby have more open conversations about online partner seeking with their clients. Past studies show equivocal results about the relationship between online partner seeking and condomless sex, with some studies among MSM from rural areas or large cities showing higher odds [30,31,86,87], whereas others report lower odds [10-12] or no associations [85,88]. Our results revealed lower odds for online partner seeking among those who reported recent condomless insertive and receptive anal sex. In a study on MSM residing outside of large urban areas, Bowen et al [71] suggested that in high-risk situations, such as meeting relatively unknown partners online, MSM tend to frequently use condoms. However, with primary partners or men to whom they feel attracted, MSM may engage in condomless sex even if they are not very familiar with their partner and are unaware of their partner's HIV status [71]. More research using relationship-level data will be valuable in better understanding how online partner seeking may play a role in affecting this association.

Evidence shows that MSM who seek partners online are more likely to report HIV [9] or STI [85] testing and willingness to use PrEP [84,89]. Consistent with previous research, we observed more (although not statistically significantly) HIV testing among users and nonusers of online tools in unadjusted analyses. MSM are recommended to be tested for HIV if they have had more than 1 sex partner since their last HIV test [90]. Therefore, the association between HIV testing and online partner seeking may be explained by increased sexual activity among users of online tools in our study and, likely, by more awareness of, and interest in, sexual health. The unadjusted analysis also revealed that a larger proportion of MSM who used online tools reported having ever taken PrEP than those who had not used online tools. However, the small number of participants who had received PrEP among nonusers of online tools prevented us from deriving consistent inferences in the adjusted analysis (ie, there were only marginally significant associations in the model for condomless receptive anal sex). In our sample, only 15.3% (36/235) of the MSM potentially eligible for PrEP use (after excluding 17 respondents who were aware of their HIV-positive status) had ever used PrEP, which is lower than the percentages for PrEP use reported from studies on MSM residing in large urban areas [91] but higher than the percentages reported in other studies on MSM from rural areas in the American South [89]. According to cross-sectionally collected data from 20 urban locations in the United States and territories as part of the National HIV Behavioral Surveillance conducted by the Centers for Disease Control and Prevention [92], there was a notable increase in PrEP awareness (from 60% to 90%) and use (from 6% to 35%) from 2014 to 2017 among MSM who are at risk for HIV infection [91]. Some authors have presumed that recent increases in PrEP awareness and use might be partly attributable to HIV prevention campaigns on social media [91]. Notably, an analysis of data from the American Men's Internet Survey [93] revealed that the likelihood of PrEP awareness and use was lower in rural and nonmetropolitan regions than in urban areas [94]. These findings in combination with our study's results indicate a need to increase PrEP awareness and uptake among nonurban MSM and highlight the need to explore social media as a strategy to do so.

Limitations

First, this was a cross-sectional survey. Second, data are subject to recall and self-report biases, although the use of an online self-administered questionnaire should mitigate this limitation because it is typically better suited for collection of sensitive data [95,96], as should the federal certificate of confidentiality which was explained to participants in the consent and survey; this provides additional protection to the data. Third, although data on antiretroviral therapy uptake were collected, data on viral load were not, leaving us unable to describe the sample in terms of viral load suppression. Fourth, this was a convenience sample of young adult MSM residing in Central Kentucky and

the sample may not be representative of all young adult MSM, especially those of lower educational attainment. Many resided in a midsized college town and were still in school or have already received their degree. A number of studies have shown that MSM who seek partners online are more likely to have higher level of education and be of higher income and, thus, are more likely to access the internet and mobile apps [9]. Therefore, our analyses were adjusted for confounding by educational attainment.

In addition, the study recruited participants through multiple sources, including but not limited to social media, leading to potential overrepresentation of men who frequently use online tools for partner seeking. A study on nonurban MSM by Bowen et al [71] and a meta-analysis by Liau et al [13] showed that more men recruited online practiced internet partner seeking than men recruited offline. Of the 252 participants in our study, 64 (25.4%) were recruited at a pride festival. Some suggest that men who attend pride festivals are more open about their sexuality and engagement in same-sex sexual activity [67]. Bowen et al [71] reported that more nonurban MSM recruited through conventional sampling methods were in long-term monogamous relationships than men recruited solely on the internet [71]. However, it is worth noting that only 20.6% (52/252) reported social media as a way in which they had learned about the survey. Some (79/252, 31.3%) of the participants reported learning about the study through more than 1 method (eg, through peer referral and seeing it on social media); hence, responses regarding the recruitment methods were not mutually exclusive. Furthermore, the proportion of MSM practicing online partner seeking did not differ significantly by the type of recruitment method, and thus we did not adjust for recruitment type in the analysis. Finally, the study did not distinguish between different types of social networking and dating websites or smartphone-based geosocial networking apps that were reported by respondents but rather focused on the overall use of various online tools for partner seeking.

Despite these limitations, our study is among the first to describe the experiences of MSM residing outside of major cities, a population that is underrepresented in research on the topic. We believe that the *lack* of differences between our findings and those of national and urban studies are as informative as the differences themselves because they provide insights into the extent to which MSM residing in small and midsized towns are similar to those residing in large cities and, therefore, may respond similarly to similar intervention approaches or experience similar challenges.

Conclusions

To our knowledge, this study is one of the first of its kind to examine online partner seeking and associated sexual behaviors among MSM from small and midsized towns and the first one in Kentucky: 71.8% (181/252) had used online tools to meet partners for sex, dating, or for both purposes. Consistent with research on MSM from larger metropolises, online partner seeking was associated with some sexual risk behaviors such as increased number of anal sex partners and substance use before or during sex. These results provide insights regarding the content of targeted HIV risk-reduction internet- or mobile-based interventions among MSM who practice online partner seeking. However, unlike most studies among MSM from rural areas or large urban centers, we observed positive associations of online partner seeking with some protective behaviors such as condom use during insertive and receptive anal intercourse. This suggests that more tailored interventions are needed to reduce the risk of HIV transmission associated with condomless anal intercourse among MSM who do not practice online partner seeking.

Acknowledgments

This study was supported by the National Institute on Drug Abuse (NIDA R03 DA039740). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. IWH is supported by the California HIV/AIDS Research Program (RP15-LA-007); the Center for HIV Identification, Prevention, and Treatment (MH58107); the University of California Los Angeles Center for AIDS Research (5P30AI028697); and the National Institutes of Health National Center for Advancing Translational Science (UL1TR000124).

Authors' Contributions

VP shaped the research question with support from AMY, conducted data analysis, interpreted the results, and drafted and revised the manuscript. AMY supervised the findings, and IWH provided critical feedback to the analysis. AMB, IWH, and AMY contributed to study design, data collection, and revisions of the manuscript. All authors approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Adjusted comparisons of users versus nonusers of online tools on sexual and drug-related behavioral characteristics with and without adjusting for metropolitan residence.
[DOCX File , 19 KB - formative_v6i6e35056_app1.docx]

References

1. HIV. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/hiv/basics/statistics.html [accessed 2021-02-02]
2. Brown G, Maycock B, Burns S. Your picture is your bait: use and meaning of cyberspace among gay men. J Sex Res 2005 Feb;42(1):63-73. [doi: 10.1080/00224490509552258] [Medline: 15795806]
3. Kirby T, Thornber-Dunwell M. Phone apps could help promote sexual health in MSM. Lancet 2014 Oct 18;384(9952):1415. [doi: 10.1016/S0140-6736(14)61849-3] [Medline: 25390311]
4. Handel M, Shklovski I. Disclosure, ambiguity and risk reduction in real-time dating sites. In: Proceedings of the 17th ACM international conference on Supporting group work. 2012 Presented at: Group '12: ACM 2012 International Conference on Support Group Work; Oct 27 - 31, 2012; Sanibel Island Florida USA. [doi: 10.1145/2389176.2389203]
5. Carballo-Diéguez A, Miner M, Dolezal C, Rosser BR, Jacoby S. Sexual negotiation, HIV-status disclosure, and sexual risk behavior among Latino men who use the internet to seek sex with other men. Arch Sex Behav 2006 Aug;35(4):473-481. [doi: 10.1007/s10508-006-9078-7] [Medline: 16933107]
6. St De Lore J, Thiede H, Cheadle A, Goldbaum G, Carey JW, Hutcheson RE, et al. HIV disclosure and subsequent sexual behaviors among men who have sex with men who meet online. J Homosex 2012 Apr;59(4):592-609. [doi: 10.1080/00918369.2012.665704] [Medline: 22500995]
7. Grov C, Agyemang L, Ventuneac A, Breslow AS. Navigating condom use and HIV status disclosure with partners met online: a qualitative pilot study with gay and bisexual men from Craigslist.org. AIDS Educ Prev 2013 Feb;25(1):72-85 [FREE Full text] [doi: 10.1521/aeap.2013.25.1.72] [Medline: 23387953]
8. Newcomb ME, Mongrella MC, Weis B, McMillen SJ, Mustanski B. Partner disclosure of PrEP use and undetectable viral load on geosocial networking apps: frequency of disclosure and decisions about condomless sex. J Acquir Immune Defic Syndr 2016 Feb 01;71(2):200-206 [FREE Full text] [doi: 10.1097/QAI.0000000000000819] [Medline: 26761520]
9. Zou H, Fan S. Characteristics of men who have sex with men who use smartphone geosocial networking applications and implications for HIV interventions: a systematic review and meta-analysis. Arch Sex Behav 2017 May;46(4):885-894. [doi: 10.1007/s10508-016-0709-3] [Medline: 27043837]
10. Rice E. Sex risk among young men who have sex with men who use Grindr, a smartphone geosocial networking application. J AIDS Clinic Res 2012;01(S4):005. [doi: 10.4172/2155-6113.S4-005]
11. Landovitz RJ, Tseng C, Weissman M, Haymer M, Mendenhall B, Rogers K, et al. Epidemiology, sexual risk behavior, and HIV prevention practices of men who have sex with men using GRINDR in Los Angeles, California. J Urban Health 2013 Aug;90(4):729-739 [FREE Full text] [doi: 10.1007/s11524-012-9766-7] [Medline: 22983721]
12. Knox J, Chen Y, He Q, Liu G, Jones J, Wang X, et al. Use of geosocial networking apps and HIV risk behavior among men who have sex with men: case-crossover study. JMIR Public Health Surveill 2021 Jan 15;7(1):e17173 [FREE Full text] [doi: 10.2196/17173] [Medline: 33448934]
13. Liau A, Millett G, Marks G. Meta-analytic examination of online sex-seeking and sexual risk behavior among men who have sex with men. Sex Transm Dis 2006 Sep;33(9):576-584. [doi: 10.1097/01.olq.0000204710.35332.c5] [Medline: 16540884]
14. Lehmiller JJ, Ioerger M. Social networking smartphone applications and sexual health outcomes among men who have sex with men. PLoS One 2014;9(1):e86603 [FREE Full text] [doi: 10.1371/journal.pone.0086603] [Medline: 24466166]
15. Young SD, Szekeres G, Coates T. The relationship between online social networking and sexual risk behaviors among men who have sex with men (MSM). PLoS One 2013;8(5):e62271 [FREE Full text] [doi: 10.1371/journal.pone.0062271] [Medline: 23658716]
16. Bolding G, Davis M, Hart G, Sherr L, Elford J. Gay men who look for sex on the internet: is there more HIV/STI risk with online partners? AIDS 2005 Jun 10;19(9):961-968. [doi: 10.1097/01.aids.0000171411.84231.f6] [Medline: 15905678]
17. Burrell ER, Pines HA, Robbie E, Coleman L, Murphy RD, Hess KL, et al. Use of the location-based social networking application GRINDR as a recruitment tool in rectal microbicide development research. AIDS Behav 2012 Oct;16(7):1816-1820 [FREE Full text] [doi: 10.1007/s10461-012-0277-z] [Medline: 22851153]
18. Grosskopf NA, LeVasseur MT, Glaser DB. Use of the internet and mobile-based "apps" for sex-seeking among men who have sex with men in New York City. Am J Mens Health 2014 Nov;8(6):510-520 [FREE Full text] [doi: 10.1177/1557988314527311] [Medline: 24658284]
19. Winetrobe H, Rice E, Bauermeister J, Petering R, Holloway IW. Associations of unprotected anal intercourse with Grindr-met partners among Grindr-using young men who have sex with men in Los Angeles. AIDS Care 2014;26(10):1303-1308. [doi: 10.1080/09540121.2014.911811] [Medline: 24754563]
20. Hospers HJ, Harterink P, Van Den Hoek K, Veenstra J. Chatters on the internet: a special target group for HIV prevention. AIDS Care 2002 Aug;14(4):539-544. [doi: 10.1080/09540120208629671] [Medline: 12204155]
21. Horvath KJ, Rosser BR, Remafedi G. Sexual risk taking among young internet-using men who have sex with men. Am J Public Health 2008 Jun;98(6):1059-1067. [doi: 10.2105/AJPH.2007.111070] [Medline: 18445804]
22. Beymer MR, Weiss RE, Bolan RK, Rudy ET, Bourque LB, Rodriguez JP, et al. Sex on demand: geosocial networking phone apps and risk of sexually transmitted infections among a cross-sectional sample of men who have sex with men in Los Angeles County. Sex Transm Infect 2014 Nov;90(7):567-572 [FREE Full text] [doi: 10.1136/sextrans-2013-051494] [Medline: 24926041]

23. Tashima KT, Alt EN, Harwell JI, Fiebich-Perez DK, Flanigan TP. Internet sex-seeking leads to acute HIV infection: a report of two cases. Int J STD AIDS 2003 Apr;14(4):285-286. [doi: 10.1258/095646203321264926] [Medline: 12716501]
24. Klausner JD, Wolf W, Fischer-Ponce L, Zolt I, Katz MH. Tracing a syphilis outbreak through cyberspace. JAMA 2000 Jul 26;284(4):447-449. [doi: 10.1001/jama.284.4.447] [Medline: 10904507]
25. Chiu CJ, Young SD. The relationship between online social network use, sexual risk behaviors, and HIV sero-status among a sample of predominately African American and Latino men who have sex with men (MSM) social media users. AIDS Behav 2015 Jun;19 Suppl 2:98-105 [FREE Full text] [doi: 10.1007/s10461-014-0986-6] [Medline: 25572831]
26. Grosskopf NA, Harris JK, Wallace BC, Nanin JE. Online sex-seeking behaviors of men who have sex with men in New York City. Am J Mens Health 2011 Sep 26;5(5):378-385 [FREE Full text] [doi: 10.1177/1557988310372801] [Medline: 20798150]
27. Patel VV, Masyukova M, Sutton D, Horvath KJ. Social media use and HIV-related risk behaviors in young Black and Latino gay and bi men and transgender individuals in New York City: implications for online interventions. J Urban Health 2016 Apr;93(2):388-399 [FREE Full text] [doi: 10.1007/s11524-016-0025-1] [Medline: 26936854]
28. Kim AA, Kent C, McFarland W, Klausner JD. Cruising on the internet highway. J Acquir Immune Defic Syndr 2001 Sep 01;28(1):89-93. [doi: 10.1097/00042560-200109010-00013] [Medline: 11579282]
29. Horvath KJ, Bowen AM, Williams ML. Virtual and physical venues as contexts for HIV risk among rural men who have sex with men. Health Psychol 2006 Mar;25(2):237-242. [doi: 10.1037/0278-6133.25.2.237] [Medline: 16569116]
30. Rosser BR, Oakes JM, Horvath KJ, Konstan JA, Danilenko GP, Peterson JL. HIV sexual risk behavior by men who use the internet to seek sex with men: results of the Men's INTernet Sex Study-II (MINTS-II). AIDS Behav 2009 Jun;13(3):488-498 [FREE Full text] [doi: 10.1007/s10461-009-9524-3] [Medline: 19205866]
31. Kakietek J, Sullivan PS, Heffelfinger JD. You've got male: internet use, rural residence, and risky sex in men who have sex with men recruited in 12 U.S. cities. AIDS Educ Prev 2011 Apr;23(2):118-127 [FREE Full text] [doi: 10.1521/aeap.2011.23.2.118] [Medline: 21517661]
32. Hubach RD, O'Neil A, Stowe M, Giano Z, Curtis B, Fisher CB. Perceived confidentiality risks of mobile technology-based ecologic momentary assessment to assess high-risk behaviors among rural men who have sex with men. Arch Sex Behav 2021 May 20;50(4):1641-1650 [FREE Full text] [doi: 10.1007/s10508-019-01612-x] [Medline: 32078710]
33. White Hughto JM, Pachankis JE, Eldahan AI, Keene DE. "You can't just walk down the street and meet someone": the intersection of social-sexual networking technology, stigma, and health among gay and bisexual men in the small city. Am J Mens Health 2017 May;11(3):726-736 [FREE Full text] [doi: 10.1177/1557988316679563] [Medline: 27885147]
34. Pennise M, Inscho R, Herpin K, Owens J, Bedard BA, Weimer AC, et al. Using smartphone apps in STD interviews to find sexual partners. Public Health Rep 2015;130(3):245-252 [FREE Full text] [doi: 10.1177/003335491513000311] [Medline: 25931628]
35. Williams ML, Bowen AM, Horvath KJ. The social/sexual environment of gay men residing in a rural frontier state: implications for the development of HIV prevention programs. J Rural Health 2005;21(1):48-55 [FREE Full text] [doi: 10.1111/j.1748-0361.2005.tb00061.x] [Medline: 15667009]
36. Preston DB, D'Augelli AR, Kassab CD, Starks MT. The relationship of stigma to the sexual risk behavior of rural men who have sex with men. AIDS Educ Prev 2007 Jun;19(3):218-230. [doi: 10.1521/aeap.2007.19.3.218] [Medline: 17563276]
37. Preston D, D'Augelli A, Cain R, Schulze F. Issues in the development of HIV-preventive interventions for men who have sex with men (MSM) in rural areas. J Primary Prevent 2002;23(2):199-214. [doi: 10.1023/A:1019968315959]
38. Preston DB, D'Augelli AR, Kassab CD, Cain RE, Schulze FW, Starks MT. The influence of stigma on the sexual risk behavior of rural men who have sex with men. AIDS Educ Prev 2004 Aug;16(4):291-303. [doi: 10.1521/aeap.16.4.291.40401] [Medline: 15342332]
39. Keene DE, Eldahan AI, White Hughto JM, Pachankis JE. 'The big ole gay express': sexual minority stigma, mobility and health in the small city. Cult Health Sex 2017 Mar;19(3):381-394 [FREE Full text] [doi: 10.1080/13691058.2016.1226386] [Medline: 27604293]
40. Annes A, Redlin M. Coming out and coming back: rural gay migration and the city. J Rural Stud 2012 Jan;28(1):56-68. [doi: 10.1016/j.jrurstud.2011.08.005]
41. Lewis N. Moving “out,” moving on: gay men's migrations through the life course. Annals Assoc Am Geographers 2014 Feb 20;104(2):225-233. [doi: 10.1080/00045608.2013.873325]
42. Algarin AB, Ward PJ, Christian WJ, Rudolph AE, Holloway IW, Young AM. Spatial distribution of partner-seeking men who have sex with men using geosocial networking apps: epidemiologic study. J Med Internet Res 2018 May 31;20(5):e173 [FREE Full text] [doi: 10.2196/jmir.9919] [Medline: 29853441]
43. HIV in the United States by region. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/hiv/statistics/overview/geographicdistribution.html [accessed 2021-02-02]
44. Rural-Urban Continuum Codes. Economic Research Service U.S. Department of Agriculture. URL: https://tinyurl.com/5h7nzymy [accessed 2021-11-01]
45. QuickFacts Fayette County, Kentucky. United States Census Bureau. URL: https://www.census.gov/quickfacts/fact/table/fayettecountykentucky/PST045219 [accessed 2021-11-01]

46. Kentucky area development district. Kentucky Atlas & Gazetteer. URL: https://www.kyatlas.com/kentucky-adds.html [accessed 2021-05-13]
47. HIV/AIDS surveillance Report 2019. Kentucky Unbridled Spirit. URL: https://chfs.ky.gov/agencies/dph/dehp/hab/Documents/AnnualReport2019.pdf [accessed 2021-05-05]
48. Van Handel MM, Rose CE, Hallisey EJ, Kolling JL, Zibbell JE, Lewis B, et al. County-level vulnerability assessment for rapid dissemination of HIV or HCV infections among persons who inject drugs, United States. J Acquir Immune Defic Syndr 2016 Nov 01;73(3):323-331 [FREE Full text] [doi: 10.1097/QAI.0000000000001098] [Medline: 27763996]
49. Health officials see increase in HIV infection among individuals who inject drugs. NKY Health. URL: https://nkyhealth.org/2018/01/09/health-officials-see-increase-in-hiv-infection-among-individuals-who-inject-drugs/ [accessed 2021-02-02]
50. As HIV cluster investigation moves into second month, health officials increase opportunities for HIV testing. NKYHEALTH. URL: https://nkyhealth.org/2018/02/22/as-hiv-cluster-investigation-moves-into-second-month-health-officials-increase-opportunities-for-hiv-testing/ [accessed 2021-02-02]
51. Number of HIV cases grows to 80 in Cabell County. WSAZ NewsChannel 3. URL: https://www.wsaz.com/content/news/HIV-cluster-confirmed-in-Cabell-County-506708841.html [accessed 2021-02-02]
52. HIV investigation. NKYHEALTH. URL: https://nkyhealth.org/individual-or-family/health-alerts/hiv-investigation/ [accessed 2022-04-11]
53. McClung RP, Atkins AD, Kilkenny M, Bernstein KT, Willenburg KS, Weimer M, 2019 Cabell County HIV Outbreak Response Team. Response to a large HIV outbreak, Cabell county, West Virginia, 2018-2019. Am J Prev Med 2021 Nov;61(5 Suppl 1):S143-S150 [FREE Full text] [doi: 10.1016/j.amepre.2021.05.039] [Medline: 34686283]
54. Ending the HIV epidemic: a plan for America. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/endhiv/docs/ending-HIV-epidemic-overview-508.pdf [accessed 2022-04-11]
55. Magnani R, Sabin K, Saidel T, Heckathorn D. Review of sampling hard-to-reach and hidden populations for HIV surveillance. AIDS 2005 May;19 Suppl 2:S67-S72. [doi: 10.1097/01.aids.0000172879.20628.e1] [Medline: 15930843]
56. Ballard AM, Cardwell T, Young AM. Fraud detection protocol for web-based research among men who have sex with men: development and descriptive evaluation. JMIR Public Health Surveill 2019 Feb 04;5(1):e12344 [FREE Full text] [doi: 10.2196/12344] [Medline: 30714944]
57. Diagnoses of HIV infection in the United States and dependent areas, 2018: gay, bisexual, and other men who have sex with men. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/hiv/library/reports/hiv-surveillance/vol-31/content/msm.html [accessed 2022-04-08]
58. Grey JA, Konstan J, Iantaffi A, Wilkerson JM, Galos D, Rosser BR. An updated protocol to detect invalid entries in an online survey of men who have sex with men (MSM): how do valid and invalid submissions compare? AIDS Behav 2015 Oct;19(10):1928-1937 [FREE Full text] [doi: 10.1007/s10461-015-1033-y] [Medline: 25805443]
59. Dewitt J, Capistrant B, Kohli N, Rosser BR, Mitteldorf D, Merengwa E, et al. Addressing participant validity in a small internet health survey (the restore study): protocol and recommendations for survey response validation. JMIR Res Protoc 2018 Apr 24;7(4):e96 [FREE Full text] [doi: 10.2196/resprot.7655] [Medline: 29691203]
60. Patel P, Borkowf CB, Brooks JT, Lasry A, Lansky A, Mermin J. Estimating per-act HIV transmission risk: a systematic review. AIDS 2014 Jun 19;28(10):1509-1519 [FREE Full text] [doi: 10.1097/QAD.0000000000000298] [Medline: 24809629]
61. Westercamp N, Mattson CL, Madonia M, Moses S, Agot K, Ndinya-Achola JO, et al. Determinants of consistent condom use vary by partner type among young men in Kisumu, Kenya: a multi-level data analysis. AIDS Behav 2010 Aug;14(4):949-959 [FREE Full text] [doi: 10.1007/s10461-008-9458-1] [Medline: 18791819]
62. Garofalo R, Herrick A, Mustanski BS, Donenberg GR. Tip of the Iceberg: young men who have sex with men, the internet, and HIV risk. Am J Public Health 2007 Jun;97(6):1113-1117. [doi: 10.2105/AJPH.2005.075630] [Medline: 17463378]
63. Celentano DD, Valleroy LA, Sifakis F, MacKellar DA, Hylton J, Thiede H, Young Men's Survey Study Group. Associations between substance use and sexual risk among very young men who have sex with men. Sex Transm Dis 2006 Apr;33(4):265-271. [doi: 10.1097/01.olq.0000187207.10992.4e] [Medline: 16434886]
64. Stueve A, O'Donnell L, Duran R, San Doval A, Geier J, Community Intervention Trial for Youth Study Team. Being high and taking sexual risks: findings from a multisite survey of urban young men who have sex with men. AIDS Educ Prev 2002 Dec;14(6):482-495. [doi: 10.1521/aeap.14.8.482.24108] [Medline: 12512849]
65. Drumright LN, Little SJ, Strathdee SA, Slymen DJ, Araneta MR, Malcarne VL, et al. Unprotected anal intercourse and substance use among men who have sex with men with recent HIV infection. J Acquir Immune Defic Syndr 2006 Nov 01;43(3):344-350. [doi: 10.1097/01.qai.0000230530.02212.86] [Medline: 16980913]
66. Newcomb ME, Clerkin EM, Mustanski B. Sensation seeking moderates the effects of alcohol and drug use prior to sex on sexual risk in young men who have sex with men. AIDS Behav 2011 Apr 20;15(3):565-575. [doi: 10.1007/s10461-010-9832-7] [Medline: 20960048]
67. Benotsch EG, Kalichman S, Cage M. Men who have met sex partners via the internet: prevalence, predictors, and implications for HIV prevention. Arch Sex Behav 2002 Apr;31(2):177-183. [doi: 10.1023/a:1014739203657] [Medline: 11974643]

68. Rosser BR, Miner MH, Bockting WO, Ross MW, Konstan J, Gurak L, et al. HIV risk and the internet: results of the Men's INTernet Sex (MINTS) study. AIDS Behav 2009 Aug;13(4):746-756 [FREE Full text] [doi: 10.1007/s10461-008-9399-8] [Medline: 18512143]
69. Garofalo R, Mustanski B, Johnson A, Emerson E. Exploring factors that underlie racial/ethnic disparities in HIV risk among young men who have sex with men. J Urban Health 2010 Mar;87(2):318-323 [FREE Full text] [doi: 10.1007/s11524-009-9430-z] [Medline: 20155329]
70. Tikkanen R, Ross MW. Technological tearoom trade: characteristics of Swedish men visiting gay internet chat rooms. AIDS Educ Prev 2003 Apr;15(2):122-132. [doi: 10.1521/aeap.15.3.122.23833] [Medline: 12739789]
71. Bowen A, Williams M, Horvath K. Using the internet to recruit rural MSM for HIV risk assessment: sampling issues. AIDS Behav 2004 Sep;8(3):311-319 [FREE Full text] [doi: 10.1023/B:AIBE.0000044078.43476.1f] [Medline: 15475678]
72. Albright JM. Sex in America online: an exploration of sex, marital status, and sexual identity in internet sex seeking and its impacts. J Sex Res 2008;45(2):175-186. [doi: 10.1080/00224490801987481] [Medline: 18569538]
73. Bauermeister JA, Leslie-Santana M, Johns MM, Pingel E, Eisenberg A. Mr. Right and Mr. Right Now: romantic and casual partner-seeking online among young men who have sex with men. AIDS Behav 2011 Feb;15(2):261-272 [FREE Full text] [doi: 10.1007/s10461-010-9834-5] [Medline: 20953689]
74. Bancroft J, Janssen E, Strong D, Carnes L, Vukadinovic Z, Long JS. Sexual risk-taking in gay men: the relevance of sexual arousability, mood, and sensation seeking. Arch Sex Behav 2003 Dec;32(6):555-572. [doi: 10.1023/a:1026041628364] [Medline: 14574099]
75. Adam PC, Teva I, de Wit JB. Balancing risk and pleasure: sexual self-control as a moderator of the influence of sexual desires on sexual risk-taking in men who have sex with men. Sex Transm Infect 2008 Nov 01;84(6):463-467. [doi: 10.1136/sti.2008.031351] [Medline: 19028948]
76. Horvath KJ, Beadnell B, Bowen AM. Sensation seeking as a moderator of internet use on sexual risk taking among men who have sex with men. Sex Res Soc Policy 2006 Dec;3(4):77-90. [doi: 10.1525/srsp.2006.3.4.77]
77. Exner TM, Meyer‐Bahlburg HF, Ehrhardt AA. Sexual self control as a mediator of high risk sexual behavior in a New York city cohort of HIV+ and HIV‐ gay men. J Sex Res 1992 Aug;29(3):389-406. [doi: 10.1080/00224499209551655]
78. Dudley MG, Rostosky SS, Korfhage BA, Zimmerman RS. Correlates of high-risk sexual behavior among young men who have sex with men. AIDS Educ Prev 2004 Aug;16(4):328-340. [doi: 10.1521/aeap.16.4.328.40397] [Medline: 15342335]
79. Nguyen HV, Koo KH, Davis KC, Otto JM, Hendershot CS, Schacht RL, et al. Risky sex: interactions among ethnicity, sexual sensation seeking, sexual inhibition, and sexual excitation. Arch Sex Behav 2012 Oct;41(5):1231-1239 [FREE Full text] [doi: 10.1007/s10508-012-9904-z] [Medline: 22350123]
80. Beymer MR, Rossi AD, Shu SB. Assessing self-control and geosocial networking app behavior among an online sample of men who have sex with men. J Urban Health 2016 Aug 16;93(4):698-708 [FREE Full text] [doi: 10.1007/s11524-016-0056-7] [Medline: 27184573]
81. Ogilvie GS, Taylor DL, Trussler T, Marchand R, Gilbert M, Moniruzzaman A, et al. Seeking sexual partners on the internet: a marker for risky sexual behaviour in men who have sex with men. Can J Public Health 2008;99(3):185-188 [FREE Full text] [Medline: 18615938]
82. Tang W, Tang S, Qin Y, Zhang Y, Zhang W, Liu C, et al. Will gay sex-seeking mobile phone applications facilitate group sex? A cross-sectional online survey among men who have sex with men in China. PLoS One 2016;11(11):e0167238 [FREE Full text] [doi: 10.1371/journal.pone.0167238] [Medline: 27880823]
83. Cultural competency, MSM and Internet-based partner services. National Coalition of STD Directors. URL: https://www.ncsddc.org/wp-content/uploads/2017/08/cultural_competency_ips_webinar_9-27-11.pdf [accessed 2020-10-09]
84. Torres TS, De Boni RB, de Vasconcellos MT, Luz PM, Hoagland B, Moreira RI, et al. Awareness of prevention strategies and willingness to use preexposure prophylaxis in Brazilian men who have sex with men using apps for sexual encounters: online cross-sectional study. JMIR Public Health Surveill 2018 Jan 22;4(1):e11 [FREE Full text] [doi: 10.2196/publichealth.8997] [Medline: 29358160]
85. Kachur RE, Copen CE, Strona FV, Bernstein KT, Furness BW, Hogben M. Factors associated with online sex partners among gay, bisexual and other men who have sex with men: results from a national survey. Sex Transm Dis 2021 Aug 01;48(8):542-546. [doi: 10.1097/OLQ.0000000000001376] [Medline: 34110750]
86. Bowen AM, Horvath K, Williams ML. A randomized control trial of internet-delivered HIV prevention targeting rural MSM. Health Educ Res 2007 Feb;22(1):120-127 [FREE Full text] [doi: 10.1093/her/cyl057] [Medline: 16849391]
87. Wang H, Zhang J, Chu Z, Hu Q, Dong W, Huang X, et al. Risk-taking behaviors and adherence to HIV pre-exposure prophylaxis in users of geosocial networking apps: real-world, multicenter study. J Med Internet Res 2020 Oct 14;22(10):e22388 [FREE Full text] [doi: 10.2196/22388] [Medline: 33052132]
88. Bien CH, Best JM, Muessig KE, Wei C, Han L, Tucker JD. Gay apps for seeking sex partners in China: implications for MSM sexual health. AIDS Behav 2015 Jun;19(6):941-946 [FREE Full text] [doi: 10.1007/s10461-014-0994-6] [Medline: 25572834]
89. Bineau L, Lambert D, Truszczynski N, Hansen N, Lauckner C. Dating app use among rural men who have sex with men and its relationship to HIV prevention and risk behaviors: a mixed-methods analysis. Rural Remote Health 2021 Jun;21(2):6556 [FREE Full text] [doi: 10.22605/RRH6556] [Medline: 34062066]

90. Getting tested. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/hiv/basics/hiv-testing/getting-tested.html [accessed 2020-10-05]
91. Finlayson T, Cha S, Xia M, Trujillo L, Denson D, Prejean J, National HIV Behavioral Surveillance Study Group. Changes in HIV preexposure prophylaxis awareness and use among men who have sex with men - 20 urban areas, 2014 and 2017. MMWR Morb Mortal Wkly Rep 2019 Jul 12;68(27):597-603 [FREE Full text] [doi: 10.15585/mmwr.mm6827a1] [Medline: 31298662]
92. National HIV behavioral surveillance (NHBS). Centers for Disease Control and Prevention. URL: https://www.cdc.gov/hiv/statistics/systems/nhbs/index.html [accessed 2021-01-21]
93. Sanchez TH, Sineath RC, Kahle EM, Tregear SJ, Sullivan PS. The annual American men's internet survey of behaviors of men who have sex with men in the United States: protocol and key indicators report 2013. JMIR Public Health Surveill 2015;1(1):e3 [FREE Full text] [doi: 10.2196/publichealth.4314] [Medline: 27227126]
94. Sullivan PS, Sanchez TH, Zlotorzynska M, Chandler CJ, Sineath RC, Kahle E, et al. National trends in HIV pre-exposure prophylaxis awareness, willingness and use among United States men who have sex with men recruited online, 2013 through 2017. J Int AIDS Soc 2020 Mar;23(3):e25461 [FREE Full text] [doi: 10.1002/jia2.25461] [Medline: 32153119]
95. Tourangeau R, Smith T. Asking sensitive questions: the impact of data collection mode, question format, and question context. Public Opinion Q 1996;60(2):275-304. [doi: 10.1086/297751]
96. Gnambs T, Kaspar K. Disclosure of sensitive behaviors across self-administered survey modes: a meta-analysis. Behav Res Methods 2015 Dec;47(4):1237-1259. [doi: 10.3758/s13428-014-0533-4] [Medline: 25410404]

Abbreviations

aOR: adjusted odds ratio
GED: General Educational Development Test
MSM: men who have sex with men
OR: odds ratio
PrEP: pre-exposure prophylaxis
PWID: persons who inject drugs
STI: sexually transmitted infection

Edited by A Mavragani; submitted 19.11.21; peer-reviewed by C Sun, V Minichiello; comments to author 25.01.22; revised version received 10.05.22; accepted 10.05.22; published 10.06.22.

Please cite as:
Pravosud V, Ballard AM, Holloway IW, Young AM
Online Partner Seeking and Sexual Behaviors Among Men Who Have Sex With Men From Small and Midsized Towns: Cross-sectional Study
JMIR Form Res 2022;6(6):e35056
URL: https://formative.jmir.org/2022/6/e35056
doi:10.2196/35056
PMID:35687395

Original Paper

A Web-Based, Provider-Driven Mobile App to Enhance Patient Care Coordination Between Dialysis Facilities and Hospitals: Development and Pilot Implementation Study

Laura C Plantinga[1], ScM, PhD; Courtney Hoge[1], MSPH; Ann E Vandenberg[1], MPH, PhD; Kyle James[1,2], MD; Tahsin Masud[1,2], MD; Anjali Khakharia[1], MBBS, MS; Carol Gray[2], MSN, DNP; Bernard G Jaar[3,4,5], MPH, MD; Janice P Lea[1,2], MD, MSCR; Christopher M O'Donnell[1,2], MD; Richard Mutell[6], MA, MBA

[1]Department of Medicine, Emory University, Atlanta, GA, United States

[2]Emory University Hospital Midtown, Emory Healthcare, Atlanta, GA, United States

[3]Department of Medicine, Johns Hopkins School of Medicine, Baltimore, MD, United States

[4]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

[5]Nephrology Center of Maryland, Baltimore, MD, United States

[6]Apex Health Innovations, Williamsburg, VA, United States

Corresponding Author:
Laura C Plantinga, ScM, PhD
Department of Medicine
Emory University
1841 Clifton Road
Room 552
Atlanta, GA, 30329
United States
Phone: 1 4047273460
Email: laura.plantinga@emory.edu

Abstract

Background: We piloted a web-based, provider-driven mobile app (*DialysisConnect*) to fill the communication and care coordination gap between hospitals and dialysis facilities.

Objective: This study aimed to describe the development and pilot implementation of *DialysisConnect*.

Methods: *DialysisConnect* was developed iteratively with focus group and user testing feedback and was made available to 120 potential users at 1 hospital (hospitalists, advanced practice providers [APPs], and care coordinators) and 4 affiliated dialysis facilities (nephrologists, APPs, nurses and nurse managers, social workers, and administrative personnel) before the start of the pilot (November 1, 2020, to May 31, 2021). Midpilot and end-of-pilot web-based surveys of potential users were also conducted. Descriptive statistics were used to describe system use patterns, ratings of multiple satisfaction items (1=not at all; 3=to a great extent), and provider-selected motivators of and barriers to using *DialysisConnect*.

Results: The pilot version of *DialysisConnect* included clinical information that was automatically uploaded from dialysis facilities, forms for entering critical admission and discharge information, and a direct communication channel. Although physicians comprised most of the potential users of *DialysisConnect*, APPs and dialysis nurses were the most active users. Activities were unevenly distributed; for example, 1 hospital-based APP recorded most of the admissions (280/309, 90.6%) among patients treated at the pilot dialysis facilities. End-of-pilot ratings of *DialysisConnect* were generally higher for users versus nonusers (eg, "I can see the potential value of *DialysisConnect* for my work with dialysis patients": mean 2.8, SD 0.4, vs mean 2.3, SD 0.6; *P*=.02). Providers most commonly selected reduced time and energy spent gathering information as a motivator (11/26, 42%) and a lack of time to use the system as a barrier (8/26, 31%) at the end of the pilot.

Conclusions: This pilot study found that APPs and nurses were most likely to engage with the system. Survey participants generally viewed the system favorably while identifying substantial barriers to its use. These results inform how best to motivate providers to use this system and similar systems and inform future pragmatic research in care coordination among this and other populations.

(JMIR Form Res 2022;6(6):e36052) doi:10.2196/36052

KEYWORDS

dialysis; hospitals; physicians; advanced practice providers; nurses; care coordination; mobile app

Introduction

Background

The coordination of care is a significant challenge in the fragmented US health care system [1]. For US patients receiving dialysis, who are hospitalized for an average of 9.5 days per year (associated with annual Medicare costs of approximately US $12 billion) [2], transitions between the outpatient dialysis facility and the hospital present unique challenges. In addition to the usual elements of successful care transitions [3], dialysis-specific issues must be coordinated with providers during hospitalization and at discharge, such as identification and mitigation of difficulty with dialysis adherence; maintenance of patients' dialysis schedules during hospitalization; and communication of changes in dry weight, vascular access status, or medications (particularly those that should be administered intravenously during outpatient dialysis treatments, such as antibiotics) to dialysis providers [4]. Although the frequency of dialysis treatment presents an opportunity to improve timely coordination for patients receiving dialysis [4], patients are likely to return to their facilities without having seen other outpatient providers who could help coordinate postdischarge care, and patients themselves may be unable to provide the reason for hospitalization or updated medical information after discharge. These care transition challenges contribute to poor outcomes, such as 30-day readmissions, which occur after approximately one-third of hospitalizations among patients receiving dialysis [2] and often before the patient presents to the dialysis facility [5]. Thus, early follow-up care after discharge is a necessary component of successful care transitions for these patients [6]: we and others have shown in previous work that time-sensitive clinical factors such as documented changes in dry weight [7], timely medication reconciliation [8], and increased physician encounters after discharge [9] are all associated with lower hospital readmission risk among patient receiving dialysis.

Electronic health records (EHRs) represent a possible means of instant information exchange between dialysis and hospital providers to address these issues; however, US outpatient dialysis facilities are usually managed independently and do not share EHRs with the hospitals to which their patients are admitted. Regional health care information exchanges are meant to circumvent these issues and close the patient information gap across health care settings [10]; however, the evidence for better outcomes is mixed [11], given that communication may not be timely, there may be no context for the information shared, and the information may not include what the provider is seeking (eg, information in medical notes rather than in the EHR). Thus, although the Centers for Medicare and Medicaid Services have explicitly prioritized the reduction of hospital readmissions in patients receiving dialysis [12-14]—and, implicitly, the improvement of care coordination between US hospitals and dialysis facilities—the lack of tools to ensure timely exchange of critical information during the hospital–dialysis facility transition remains a primary barrier to improving hospital outcomes in this population [15].

Objective

To address this barrier, we conducted a pragmatic pilot study in which we developed *DialysisConnect*, a secure web-based provider communication platform, and implemented it at our 4 Emory Dialysis outpatient facilities and Emory University Hospital Midtown (EUHM). Emory Dialysis and EUHM share an academic affiliation but do not share health care management or, importantly, an EHR. We aimed to (1) gather information on the critical components of the system from stakeholders and potential users, including hospital providers (hospitalists, advanced practice providers [APPs; nurse practitioners or physician assistants], and social workers or care coordinators), and dialysis providers (nephrologists, APPs, nurses, and social workers); (2) develop and introduce the system into the clinical setting with the help of site champions to determine which potential users engaged with the system and how; and (3) examine provider perceptions of the system. Together, this information on the feasibility, acceptability, and potential sustainability of this system, as piloted, is important for future research addressing communication gaps and care coordination between settings for patients receiving dialysis. Here, we describe the development and pilot implementation of *DialysisConnect*.

Methods

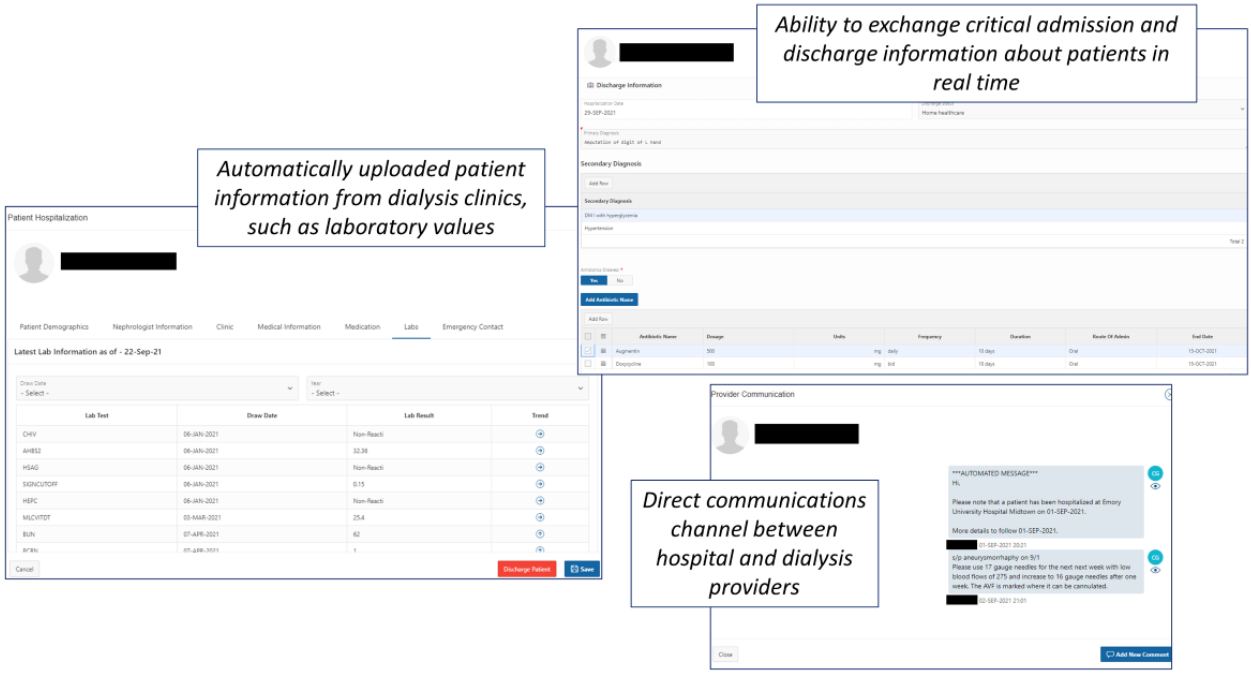
Development of DialysisConnect

The initial wireframes for *DialysisConnect*, which were created by the technical team (led by RM) and showed each intended task (admission, communication, and discharge), were based on an existing platform for the exchange of transplant referral information for patients receiving dialysis [16]. The proposed workflow process for *DialysisConnect* was developed by the research team (including experts in epidemiology and health services research, hospital medicine, nephrology, gerontology, and engineering) [17]. Initial feedback on this *DialysisConnect* prototype (video simulation using static wireframes) was collected from 4 focus groups of stakeholders and potential users of *DialysisConnect* (excluding the site champions), as previously described [17]. The purpose of the focus groups (led by AEV) was to gather feedback on the desired elements and features of the system and increase future buy-in by the inclusion of potential users of *DialysisConnect* in its creation. Data regarding feedback on the proposed system were collated from the transcripts for the technical team (Multimedia Appendix 1). Incorporation of this initial feedback into the test system was prioritized by its importance to the focus group participants, the degree of automation that was possible, and the ease of building features into the system over a short period. Research team members thoroughly evaluated this test system and presented

it to the clinical teams at the hospital and dialysis facilities in February 2020; disruptions to clinical care early in the COVID-19 pandemic led to an 8-month delay in the rollout, during which additional planned future upgrades to the system and user testing (n=9 potential users, including the *site champions* [JPL and TM at Emory Dialysis and CMO and KJ at EUHM]) were performed. Each participant was asked to perform a standardized set of tasks on the test system, with guidance from the research team (LCP and CH). Observations and feedback (Multimedia Appendix 2) were provided to the technical team for a final round of fixes and updates to the system before rollout.

Just before rollout, the team conducted remote group training sessions using the pilot version of *DialysisConnect* (Figure 1; Multimedia Appendix 3), and group emails were sent to all potential *DialysisConnect* users (as identified by champions) to announce the live rollout, along with detailed user guides (Multimedia Appendices 4 and 5) and contact information for questions. Throughout the pilot study, the list of potential users was updated as needed, and all significant upgrades (Textbox 1) were announced via email, including reminders about the availability of on-demand training sessions and revised user guides (Multimedia Appendices 4 and 5). The user guides included an idealized scenario of use for both hospital and dialysis providers, in which (1) a hospital provider identified an Emory Dialysis patient at admission; (2) the hospital provider reviewed any relevant, automatically uploaded information about the patient (eg, nephrologist, clinic, recent laboratories, and medications) and checked the reasons for admission; (3) a dialysis provider received an automated message about admission generated by the hospital provider's entry and logged in to review the patient and status; (4) hospital and dialysis providers communicated via instant messages as needed or desired about the patient's status throughout the hospitalization; (5) a hospital provider clicked through the discharge elements at the patient's discharge; and (6) dialysis providers reviewed the discharge elements and, as needed, confirmed receipt (eg, for antibiotic orders). Given that this was a pilot in which we hoped to learn how staff engaged with the system, no specific expectations or benchmarks for the use of the system were communicated by the research team, although site champions were free to encourage use through whatever means they preferred. *DialysisConnect* remains available to all users; however, there has been no active encouragement of the use of *DialysisConnect* since May 2021.

Figure 1. Overview of the major features of the DialysisConnect system.



Textbox 1. Substantial upgrades made to DialysisConnect during the pilot study.

Upgrades

- November 19, 2020: automated file feed from Emory Dialysis updated to include nephrologist name and emergency contact information
- January 29, 2021: automated file feed from Emory Dialysis updated to include laboratory values and medications; option for users to view graphs of laboratory values over time added
- April 14, 2021: addition of "response required" option on messages (vs "read" status only), so that recipient was required to acknowledge the sender's message (eg, to acknowledge antibiotic orders were received and acted upon)

Pilot Implementation of DialysisConnect

Data Sources

Evaluation of the pilot implementation of *DialysisConnect* was primarily based on the system and provider survey data. System data downloaded directly from *DialysisConnect* included data on users, hospitalizations, page activities, and exchanged messages. Users were defined as potential users (all providers who had access to *DialysisConnect*), active users (logged in at least once during the pilot), and top users (defined by the ≥90th percentile of the number of log-ins or ≥45 log-ins). Daily user and activity data were aggregated by study week. Brief web-based surveys were sent to current potential users (regardless of actual use) at both Emory Dialysis and EUHM in January 2021 (n=106; midpilot) and in May 2021 (n=116; end of pilot); midpilot and end-of-pilot surveys included identical items related to actions and beliefs regarding *DialysisConnect*, adapted from implementation measures that were guided by the normalization process theory [18,19] and scored on a Likert scale (1=not at all; 2=to some extent; 3=to a great extent), as well as items assessing motivators of and barriers to using the system (Multimedia Appendix 6). Finally, EHR data from EUHM were used to identify hospitalizations that had occurred from November 1, 2020, to May 31, 2021, among Emory Dialysis patients.

Analysis

The system and user survey data were summarized using descriptive statistics. For survey data, the overall characteristics of participants and participant-identified motivators of and barriers to the use of *DialysisConnect* were described, and ratings of items related to actions and beliefs regarding *DialysisConnect* were described overall and stratified by user status and setting (dialysis facility vs hospital) at the time of the survey; paired 2-tailed *t* tests were used to compare midpilot and end-of-pilot survey ratings for the users who responded to both surveys. The capture of hospitalizations in *DialysisConnect* was estimated as the percentage of hospitalizations documented in the EHR among Emory Dialysis patients who were entered into *DialysisConnect* over the course of the pilot. All analyses were performed using Stata (version 17.0; StataCorp).

Ethical Considerations

Provider participants in the initial focus groups and midpilot and end-of-pilot surveys provided informed consent and were incentivized with meals (focus groups) and nominal (US $10) gift cards (surveys). In this pragmatic study, potential users in the pilot were not incentivized to use the system. The need for consent was waived for the EHR and system data, which were only reported in aggregated form. The entire study protocol was approved by the Emory Institutional Review Board (IRB00102971).

Results

Features of DialysisConnect

An overview of the features of the pilot version of *DialysisConnect* is presented in Figure 1. The home page included an overall hospitalization report, listings of patients who were hospitalized at the time of the study and were previously hospitalized, and a message center (Multimedia Appendix 3). The admission feature allowed hospital providers to search for and select patients and review medical information about the patients and identify reasons for hospitalization; instant automated messages informed dialysis providers of admissions (Multimedia Appendix 3). The message feature allowed providers to view automated messages (at admission, discharge, and document upload), send and view user-initiated messages, check *read* status for messages, and request and upload documents (Multimedia Appendix 3). To prompt log-ins, users also received automated messages externally, by email (default) or SMS text messages, with links to the system. Hospital providers could use the discharge feature to complete a brief, instantaneously delivered discharge report, including discharge date, status, and diagnosis; antibiotics to order; medication changes; and dialysis prescription or dry weight changes (Multimedia Appendix 3). The system was flexible and allowed modifications during the pilot based on user requests or inputs (Textbox 1).

DialysisConnect Users and Activity

Characteristics of Users

A total of 120 individuals at EUHM and Emory Dialysis (identified by project champions) had access to *DialysisConnect* over the course of the pilot (n=61, 50.8% from EUHM and n=59, 49.2% from Emory Dialysis; Figure 1). Potential hospital users were primarily hospitalists but also included APPs, nephrology fellows, and care coordinators and social workers. Most potential dialysis facility users were nurses or nurse managers, followed by nephrologists, APPs, vascular access team members, and a dietitian (Table 1).

Table 1. Cumulative *DialysisConnect* user activity and message activity during the pilot (November 1, 2020 to May 31, 2021)—overall and by role.

User role	Potential users (n=120), n (%)	Active users[a] (n=46), n (%)	Top users[b] (n=11), n (%)	Total log-ins (n=1699), n (%)	Total messages[c] (n=1145), n (%)	User-initiated messages only (n=573), n (%)
Hospital						
All hospital users	61 (100)	16 (100)	1 (100)	300 (100)	913 (100)	348 (100)
Hospitalist	49 (80.3)	9 (56.3)	0 (0)	21 (7)	7 (0.8)	5 (1.4)
APP[d]	4 (6.6)	3 (18.8)	1 (100)	238 (79.3)	861 (94.3)	339 (97.4)
Nephrology fellow	5 (8.2)	4 (25)	0 (0)	41 (13.7)	45 (4.9)	4 (1.1)
Care coordinator or social worker	3 (4.9)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Dialysis facility						
All dialysis facility users	59 (100)	30 (100)	10 (100)	1399 (100)	232 (100)	225 (100)
Nephrologist	12 (20.3)	7 (23.3)	0 (0)	54 (3.9)	7 (3)	7 (3.1)
APP	2 (3.4)	2 (6.7)	2 (20)	277 (19.8)	202 (87.1)	202 (89.8)
Dialysis nurse or nurse manager	26 (44.1)	21 (70)	8 (80)	1068 (76.3)	23 (9.9)	16 (7.1)
Dietitian	1 (1.7)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Vascular access team	2 (3.4)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Administrative assistant	7 (11.9)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Social worker	9 (15.3)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)

[a]Logged into the system at least once over the course of the pilot.

[b]Top users were in the ≥90th percentile for all potential users (≥45 log-ins).

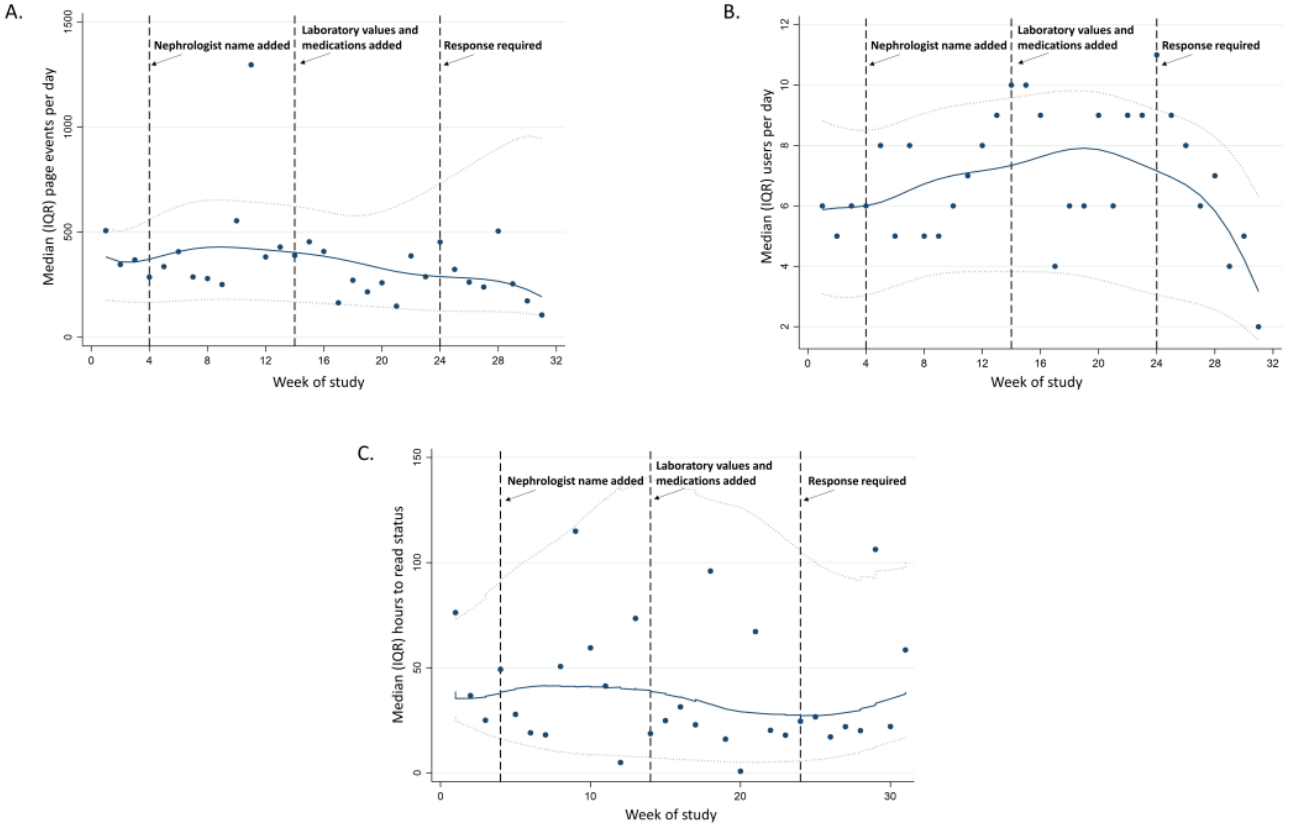
[c]Includes automated messages sent by the system at admission, at discharge, and when documents were uploaded.

[d]APP: advanced practice provider.

DialysisConnect Log-Ins

From November 1, 2020, to May 31, 2021, of the 120 potential users included in the analysis, 46 (38.3%) were active users and 11 (9.1%) were top users (Table 1). Hospital APPs (238/300, 79.3% log-ins; 217/238, 91.2% by a single APP [CG]), dialysis facility APPs (277/1399, 19.80% log-ins), and dialysis facility nurses and nurse managers (1068/1399, 76.34% log-ins) were responsible for most of the system log-ins (Table 1). Physicians were responsible for 4.41% (75/1699) of log-ins; care coordinators (hospital) and the dietitian, vascular access team members, administrative assistants, and social workers (dialysis facility) never logged in to *DialysisConnect* (Table 1). Over the course of the pilot, activity, as measured by the number of page events (number of pages with which users interacted, eg, home page and admission page) and unique users logging in (Figure 2) per week, was relatively stable, with no evidence of increase at the time of substantial upgrades (Textbox 1), and there was a decline in both measures at the end of the pilot.

Figure 2. The (A) median number of DialysisConnect page events (number of pages with which users interacted; eg, home page and admission page), (B) number of DialysisConnect users, and (C) the time (hours) from when the message was sent to when the message was read per week over the period from November 1, 2020, to May 31, 2021.



DialysisConnect Messages

Table 1 shows the total message activity in *DialysisConnect* during the pilot. A total of 1145 messages were sent, 573 (50.04%) of which were user-initiated messages (541/573, 94.4% by a single hospital APP [CG] and dialysis facility APPs; Table 1). Over the course of the pilot, the median number of hours to *read* status for messages sent varied but was <48 hours in all weeks of the pilot (Figure 2). All 21 messages that were sent with the *response required* option (April 14, 2021, to May 31, 2021) received a response, with a median response time of 18.3 (IQR 2.3-70.9) hours.

Hospitalization Events Entered Into DialysisConnect

A total of 309 incident hospitalization events, representing 184 unique patients, were entered into *DialysisConnect* during the study period by a single hospital APP (CG; 280/309, 90.6%) and 2 nephrology fellows (29/309, 9.4%). Most events (276/309, 89.3%) were among patients receiving hemodialysis, and the remaining events were among patients treated with peritoneal dialysis (31/309, 10%) or an unspecified modality (2/309, 0.6%). Of the 309 hospital events, 296 (95.8%; 178/184, 96.7% of individuals) were among active Emory Dialysis patients whose events occurred in the period from November 1, 2020, to May 31, 2021, and were included in the EUHM EHR (n=296, 81.3% of all 364 hospital events recorded for Emory Dialysis patients in the EUHM EHR in the same period). The capture of events was higher for inpatient admissions (223/260, 85.8%) than for observational stays (73/104, 70.2%). After 11 months from the end of the pilot (April 30, 2022), 335 additional hospitalization events were entered into the system.

User Perceptions of DialysisConnect

Characteristics of Survey Participants

Midpilot and end-of-pilot surveys were completed by 21.7% (23/106) and 22.4% (26/116) of potential users; 23% (9/40) of respondents completed both surveys. Participants were primarily physicians, APPs, and nurses who had been working in their setting for >12 months. Most were actual users at the time of the survey, although this dropped from 68% (15/22) to 54% (14/26) by the end of the pilot (Table 2).

Table 2. Characteristics of *DialysisConnect* user survey participants (N=40[a]).

Characteristic	Survey administered, n (%)	
	Midpilot (n=23)	End of pilot (n=26)
Site[b]		
Dialysis facility	10 (43)	15 (58)
Hospital	13 (57)	11 (42)
Role		
Physician	11 (48)	10 (39)
APP[c]	5 (22)	4 (15)
Nurse	3 (13)	7 (27)
Social worker	1 (4)	2 (8)
Other	3 (13)	3 (12)
Length of time in role (months)		
<6	3 (13)	1 (4)
6-12	3 (13)	0 (0)
>12	17 (74)	25 (96)
User of the system at time of survey		
Yes	15 (68)	14 (54)
No	7 (32)	12 (46)

[a]Nine participants filled out both surveys, leaving 40 unique individuals across both surveys.

[b]Nephrologists included in the dialysis facility group.

[c]APP: advanced practice provider.

Participant Ratings of DialysisConnect

Overall, the mean ratings of the items representing actions and beliefs about *DialysisConnect* were positive (Table 3). For both the midpilot and end-of-pilot surveys, users rated items more positively than nonusers; for example, end-of-pilot ratings were 2.8 versus 2.3 (*P*=.02) for "I can see the potential value of DialysisConnect for my work with dialysis patients" and 2.4 versus 1.8 (*P*=.04) for "Sufficient training is provided to enable staff to implement DialysisConnect" (Table 3). In general, participants working in dialysis facilities provided more positive ratings than those working in the hospital, although they rated the system as more disruptive to working relationships (end-of-pilot ratings, 2.5 vs 1.9; *P*=.03); only a few of the differences between participants in the 2 settings were statistically significant (Table 4). For the participants who completed both surveys (9/40, 23%), there were no differences between the midpilot and end-of-pilot survey ratings, except for the management support of *DialysisConnect*, which, on average, was rated more positively at midpilot than at the end of pilot (2.9 vs 2.3), although this difference was not statistically significant (*P*=.10).

Table 3. Ratings of *DialysisConnect* by survey participants in midpilot (1/21) and end-of-pilot (5/21) surveys—overall and by user status—at the time of the survey.

Item	Rating (1=not at all; 2=to some extent; 3=to a great extent)							
	Midpilot survey				End-of-pilot survey			
	Overall[a], mean (SD)	Users[a], mean (SD)	Nonusers[a], mean (SD)	*P* value[b]	Overall, mean (SD)	Users[a], mean (SD)	Nonusers[a], mean (SD)	*P* value[b]
I can see how *DialysisConnect* differs from usual ways of communicating with dialysis facilities or hospitals	2.2 (0.5)	2.3 (0.6)	2.0 (0.0)	.17	2.5 (0.6)	2.6 (0.5)	2.3 (0.7)	.18
I understand how *DialysisConnect* could improve transitions of care for dialysis patients	2.3 (0.5)	2.4 (0.5)	2.1 (0.4)	.25	2.6 (0.6)	2.8 (0.4)	2.3 (0.7)	.04
I can see the potential value of *DialysisConnect* for my work with dialysis patients	2.4 (0.6)	2.5 (0.6)	2.1 (0.4)	.23	2.5 (0.6)	2.8 (0.4)	2.3 (0.6)	.02
Users of the system in this organization have a shared understanding of the purpose of *DialysisConnect*	2.4 (0.5)	2.5 (0.5)	2.2 (0.4)	.14	2.3 (0.6)	2.3 (0.7)	2.3 (0.5)	.76
There are key people at my institution who drive *DialysisConnect* forward and get others involved	2.5 (0.6)	2.6 (0.5)	2.2 (0.8)	.11	2.3 (0.6)	2.4 (0.8)	2.1 (0.4)	.30
I believe that participating in *DialysisConnect* is a legitimate part of my role in caring for dialysis patients	2.3 (0.7)	2.4 (0.7)	2.0 (0.6)	.22	2.5 (0.6)	2.7 (0.6)	2.2 (0.6)	.03
I am open to working with colleagues to optimize our use of *DialysisConnect* for patient care	2.6 (0.5)	2.7 (0.5)	2.3 (0.5)	.05	2.5 (0.6)	2.6 (0.6)	2.3 (0.5)	.18
I support *DialysisConnect*	2.5 (0.5)	2.7 (0.5)	2.1 (0.4)	.02	2.6 (0.5)	2.8 (0.4)	2.3 (0.5)	.02
I can easily integrate *DialysisConnect* into my existing work	2.1 (0.7)	2.3 (0.7)	1.7 (0.5)	.08	2.3 (0.8)	2.5 (0.9)	2.2 (0.7)	.30
DialysisConnect disrupts working relationships[c]	1.3 (0.7)	1.4 (0.8)	1.2 (0.4)	.48	1.3 (0.6)	1.2 (0.6)	1.5 (0.5)	.27
I have confidence in other people's ability to use *DialysisConnect*	2.2 (0.5)	2.3 (0.5)	2.0 (0.6)	.25	2.2 (0.7)	2.3 (0.7)	2.2 (0.6)	.66
Work is assigned to those with skills appropriate to *DialysisConnect*	2.2 (0.7)	2.3 (0.8)	2.0 (0.0)	.57	2.1 (0.6)	2.1 (0.8)	2.0 (0.0)	.80
Sufficient training is provided to enable staff to implement *DialysisConnect*	2.2 (0.4)	2.3 (0.5)	2.0 (0.0)	.25	2.1 (0.7)	2.4 (0.7)	1.8 (0.5)	.04
Sufficient resources are available to support *DialysisConnect*	2.3 (0.5)	2.4 (0.5)	2.0 (0.0)	.16	2.4 (0.6)	2.5 (0.7)	2.1 (0.4)	.21
Management adequately supports *DialysisConnect*	2.5 (0.5)	2.6 (0.5)	2.0 (0.0)	.03	2.3 (0.7)	2.5 (0.7)	2.0 (0.5)	.10
I value the effects that *DialysisConnect* has had on my work	2.2 (0.9)	2.4 (0.8)	1.5 (0.6)	.08	2.4 (0.7)	2.5 (0.7)	2.0 (0.6)	.11

Item	Rating (1=not at all; 2=to some extent; 3=to a great extent)							
	Midpilot survey				End-of-pilot survey			
	Overall[a], mean (SD)	Users[a], mean (SD)	Nonusers[a], mean (SD)	*P* value[b]	Overall, mean (SD)	Users[a], mean (SD)	Nonusers[a], mean (SD)	*P* value[b]
The staff here agree that *DialysisConnect* is worthwhile	2.3 (0.6)	2.4 (0.5)	2.0 (0.7)	.19	2.3 (0.5)	2.5 (0.5)	2.0 (0.0)	.06
Feedback about *DialysisConnect* can be used to improve it in the future	2.5 (0.5)	2.7 (0.5)	2.3 (0.5)	.10	2.5 (0.5)	2.6 (0.5)	2.4 (0.5)	.27
I can easily modify how I work with *DialysisConnect*	2.1 (0.7)	2.3 (0.7)	1.7 (0.5)	.05	2.2 (0.7)	2.4 (0.8)	2.0 (0.4)	.20

[a]N=22 (n=15, 68% users and n=7, 32% nonusers) for the midpilot survey; N=26 (n=14, 54% users and n=12, 46% nonusers) for the end-of-pilot survey. One of the respondents did not answer the question regarding system use.

[b]Users versus nonusers at the time of the survey by *t* test.

[c]Ratings were flipped for this item (1=not at all disruptive; 3=disruptive to a great extent).

Table 4. Ratings of *DialysisConnect* by survey participants in midpilot (1/21) and end-of-pilot (5/21) surveys—overall and by participant setting.

Item	Rating (1=not at all; 2=to some extent; 3=to a great extent)							
	Midpilot survey				End-of-pilot survey			
	Overall[a], mean (SD)	Hospital[a], mean (SD)	Dialysis facility[a], mean (SD)	*P* value[b]	Overall[a], mean (SD)	Hospital[a], mean (SD)	Dialysis facility[a], mean (SD)	*P* value[b]
I can see how *DialysisConnect* differs from usual ways of communicating with dialysis facilities or hospitals	2.2 (0.5)	2.2 (0.4)	2.3 (0.7)	.52	2.5 (0.6)	2.3 (0.6)	2.7 (0.5)	.09
I understand how *DialysisConnect* could improve transitions of care for dialysis patients	2.3 (0.5)	2.2 (0.4)	2.5 (0.5)	.20	2.6 (0.6)	2.3 (0.6)	2.8 (0.4)	.02
I can see the potential value of *DialysisConnect* for my work with dialysis patients	2.4 (0.6)	2.2 (0.6)	2.6 (0.5)	.14	2.5 (0.6)	2.2 (0.6)	2.8 (0.4)	.005
Users of the system in this organization have a shared understanding of the purpose of *DialysisConnect*	2.5 (0.5)	2.4 (0.5)	2.6 (0.5)	.42	2.3 (0.6)	2.0 (0.5)	2.5 (0.5)	.05
There are key people at my institution who drive *DialysisConnect* forward and get others involved	2.5 (0.6)	2.5 (0.5)	2.6 (0.7)	.84	2.3 (0.6)	2.2 (0.7)	2.4 (0.7)	.58
I believe that participating in *DialysisConnect* is a legitimate part of my role in caring for dialysis patients	2.3 (0.7)	2.2 (0.7)	2.5 (0.7)	.25	2.5 (0.6)	2.1 (0.7)	2.7 (0.5)	.01
I am open to working with colleagues to optimize our use of *DialysisConnect* for patient care	2.6 (0.5)	2.5 (0.5)	2.7 (0.5)	.45	2.5 (0.6)	2.3 (0.6)	2.7 (0.5)	.09
I support *DialysisConnect*	2.5 (0.5)	2.5 (0.5)	2.5 (0.5)	.86	2.6 (0.5)	2.4 (0.5)	2.7 (0.5)	.06
I can easily integrate *DialysisConnect* into my existing work	2.1 (0.7)	2.0 (0.7)	2.2 (0.6)	.49	2.3 (0.8)	1.9 (0.8)	2.7 (0.6)	.01
DialysisConnect disrupts working relationships[c]	1.3 (0.7)	1.2 (0.6)	1.5 (0.8)	.32	1.3 (0.6)	1.3 (0.5)	1.3 (0.6)	.92
I have confidence in other people's ability to use *DialysisConnect*	2.2 (0.5)	2.1 (0.5)	2.4 (0.5)	.14	2.2 (0.7)	1.9 (0.5)	2.5 (0.6)	.03
Work is assigned to those with skills appropriate to *DialysisConnect*	2.2 (0.7)	2.1 (0.8)	2.3 (0.5)	.55	2.1 (0.6)	1.9 (0.6)	2.2 (0.6)	.30
Sufficient training is provided to enable staff to implement *DialysisConnect*	2.3 (0.5)	2.3 (0.5)	2.3 (0.5)	.87	2.1 (0.7)	2.1 (0.6)	2.1 (0.7)	.91
Sufficient resources are available to support *DialysisConnect*	2.3 (0.5)	2.3 (0.5)	2.4 (0.5)	.76	2.4 (0.6)	2.3 (0.5)	2.4 (0.7)	.81
Management adequately supports *DialysisConnect*	2.5 (0.5)	2.3 (0.5)	2.7 (0.5)	.18	2.3 (0.7)	2.0 (0.8)	2.5 (0.5)	.10
I value the effects that *DialysisConnect* has had on my work	2.2 (0.8)	2.0 (0.9)	2.4 (0.7)	.35	2.4 (0.7)	2.1 (0.8)	2.5 (0.5)	.19

Item	Rating (1=not at all; 2=to some extent; 3=to a great extent)							
	Midpilot survey				End-of-pilot survey			
	Overall[a], mean (SD)	Hospital[a], mean (SD)	Dialysis facility[a], mean (SD)	*P* value[b]	Overall[a], mean (SD)	Hospital[a], mean (SD)	Dialysis facility[a], mean (SD)	*P* value[b]
The staff here agree that *DialysisConnect* is worthwhile	2.3 (0.6)	2.3 (0.6)	2.3 (0.5)	.78	2.3 (0.5)	2.3 (0.5)	2.4 (0.5)	.74
Feedback about *DialysisConnect* can be used to improve it in the future	2.6 (0.5)	2.6 (0.5)	2.5 (0.5)	.60	2.5 (0.5)	2.4 (0.5)	2.7 (0.5)	.14
I can easily modify how I work with *DialysisConnect*	2.2 (0.7)	2.1 (0.8)	2.3 (0.7)	.47	2.2 (0.7)	2.0 (0.8)	2.3 (0.6)	.23

[a]N=23 (n=13, 57% hospital and n=10, 43% dialysis facilities) for the midpilot survey; N=26 (n=11, 42% users and n=15, 58% nonusers) for the end-of-pilot survey.

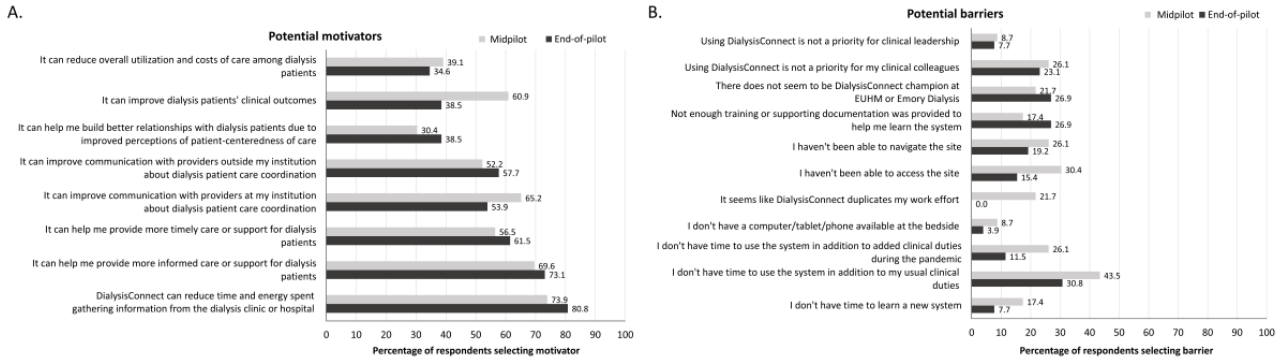
[b]Participants from the hospital versus dialysis facilities at the time of the survey by *t* test.

[c]Ratings were flipped for this item (1=not at all disruptive; 3=disruptive to a great extent).

User-Identified Motivators of and Barriers to Using DialysisConnect

Overall, among the lists of potential motivators and barriers, motivators (Figure 3) were more commonly selected by the survey participants than barriers (Figure 3). The most common motivators selected were the reduction in time spent gathering information, more informed care, and more timely care and support (21/26, 81%; 19/26, 73%; and 16/26, 62%, respectively, at the end of the pilot; Figure 3). The most common barriers selected were lack of time in addition to usual duties, lack of site champions, and insufficient training or documentation (8/26, 31%; 7/26, 27%; and 7/26, 27%, respectively, at the end of the pilot; Figure 3). When asked to rank the top motivator, participants most commonly selected the potential to gather information, the ability to provide more informed care, and the potential to improve communication with outside providers (midpilot: 6/23, 26%; 6/23, 26%; and 4/23, 17%, respectively; end of pilot: 11/26, 42%; 4/26, 15%; and 4/26, 15%, respectively). At midpilot, the top-ranked barriers were lack of time in addition to usual duties (4/23, 17%) and insufficient training or documentation, inability to navigate the site, and perceived duplication of work (3/23, 13% each); at the end of the pilot, lack of time, in addition to usual duties (8/26, 31%) and inability to navigate the site (5/26, 19%), were the most common top-ranked barriers, and no participants named perceived duplication of work as a top barrier to the use of *DialysisConnect*.

Figure 3. (A) Motivators of and (B) barriers to the use of DialysisConnect identified by user survey participants in the user survey. Participants could select >1 motivator and barrier; hence, total percentages exceed 100%. EUHM: Emory University Hospital Midtown.



Other Participant Feedback

By the end of the pilot, 86% (12/14) of those using *DialysisConnect* reported intending to keep using the system, and 92% (11/12) of those who did not use *DialysisConnect* reported intending to start. When asked for suggestions for improvement, participants suggested features that were not included (eg, including patients admitted to non-Emory hospitals and integration of *DialysisConnect* into the EHR); however, they also listed multiple features and support that were provided either throughout the pilot (eg, web-based training) or during the pilot in response to real-time feedback (eg, updating laboratories and medications and sending notifications to physicians only about their patients). In free-text responses, participants identified issues with suboptimal use ("...system functions well if staff will use it") and internal communication about the system ("...don't recall any of the clinical admin or patient care staff mentioning DialysisConnect...") but also praised the system's utility ("I believe it to be excellent source to improve communication and care for our patients" and "It has been and continues to be a great help in caring for the in-center dialysis patients...") and ease of use ("...system is user-friendly and even on the busiest days doesn't take provider more than a few minutes to add or update a patient...").

Discussion

Principal Findings

In this study, we developed and piloted a flexible web-based communication portal between providers at dialysis facilities and hospitals, the design and content of which were driven by potential users in both settings, suggesting that *DialysisConnect* is a potentially feasible solution to the gap in communication between dialysis facilities and hospitals. Although most of the potential users with access to the system were physicians, APPs in both settings and dialysis nurses were the most active users during the pilot. Our data also suggest that *DialysisConnect* is a usable solution. In midpilot and end-of-pilot surveys, both users and nonusers of the system rated *DialysisConnect* positively, and users also stated that the system was easy to use and did not disrupt their workflow. Survey participants identified multiple motivators for using *DialysisConnect*, predominantly the potential of the system to reduce time and energy spent in gathering information from other settings. However, future work in this area would need to address several reported barriers to the use of *DialysisConnect*, including a lack of time in addition to usual clinical duties, perceived lack of a system champion, and insufficient training and documentation.

Provider perceptions of lack of time in addition to usual clinical duties was an expected barrier in this pilot, given the new, untested nature of the system and the general challenges of affecting provider behavior change. Including multiple, active, and contextually relevant behavior change strategies, such as continuing education incentives, might help overcome this barrier [20,21]. In fact, we found that the barrier of perceived duplicated work effort was no longer identified as a barrier by participants at the end of the pilot study, suggesting that users may have dropped processes made obsolete by *DialysisConnect* (eg, phone calls).

Providers also commonly reported that a lack of training and supporting documentation was a barrier to using *DialysisConnect*. However, our research team provided multiple group and individual training sessions and detailed user guides for potential users in both settings at the start of the pilot, as well as additional, as-needed group or individual training sessions and updated user guides with every *DialysisConnect* user communication. The perception of lack of training and documentation despite these efforts may have been partially because of COVID-19–related disruptions; the research team was unable to provide formal and informal in-person training, as originally planned, and many potential users may not have had web-based meetings and email communications in their usual workflow and, thus, may have missed training opportunities. Importantly, the perception of lack of training is also likely partially because of another commonly reported barrier: the reported lack of support by management and key people driving *DialysisConnect* forward. This lack of support may have driven poor attendance in the web-based group training sessions.

In addition to the abovementioned barriers, several potential users noted in the surveys that the lack of integration, or even an icon to enter *DialysisConnect* from the desktop rather than navigating to the website, was a barrier. The lack of EHR integration also meant that automated information on patients had to be uploaded via patient census files that were provided by the dialysis facilities (but not hospitals), causing some lags in the information, particularly for new patients. The feasibility of including this information for other dialysis facilities or hospitals in future implementation would depend on the willingness of health care administration and information technology to provide such files or facilitate EHR integration of the system. However, the lack of EHR integration provided some advantages as well, as the Health Insurance Portability and Accountability Act–compliant system was web-based and, therefore, EHR-agnostic. It was implementable in both settings with fewer institution-specific, often burdensome information technology compliance requirements than EHR-integrated systems. Thus, we were able to offer the system to potential users across both settings and make changes to the system in real time.

Despite these barriers, 81.3% (296/364) of hospital admissions were captured over the course of the pilot study. Furthermore, this capture rate was accomplished by a single APP in the hospital setting and suggests that fidelity to the intervention can be high, even with few users: overall, a small number of users, who were primarily APPs and nurses, were responsible for most *DialysisConnect* activities. APPs were far more active than physicians at either site, which may reflect the APPs' greater time available for clinical care, site champion encouragement of APPs specifically, and the existing roles of APPs in institutions as independent physician extenders. In fact, our APPs were already performing most of the intersetting communication at the start of the pilot. Although social workers expressed initial enthusiasm for *DialysisConnect* in early focus groups [17], none of the social workers, who were given full access to the system and who facilitated critical nonclinical support (eg, transportation and financial assistance) to ensure a successful care transition, logged into the system during the pilot. This may reflect a lack of site champion encouragement of social workers; the perception of *DialysisConnect* as a platform to exchange medical information only; or particular disengagement of social workers, given pandemic-related increased workloads (and, for dialysis social workers, adjustments because of working from home during our pilot). Other essential care transition roles, such as care coordinators and vascular access team members, were also not represented among the users.

Finally, there is evidence that *DialysisConnect* could be a sustainable intervention. Most survey participants who were users of *DialysisConnect* stated that they would continue using the system at the end of the pilot. Furthermore, we found that many hospitalization events continued to be entered into *DialysisConnect* after the conclusion of the pilot when no communication encouraging its use was being sent by the research team. Site champions who identify and assign users at defined points in the workflow are likely to increase their sustainability. Sustainability is a key attribute for interventions such as *DialysisConnect*, which would be used in pragmatic settings where research teams are usually not embedded.

Comparison With Prior Work

The reporting of interventions to reduce readmissions, specifically in the population of patients receiving dialysis, is limited to a few quality improvement projects and is generally focused on effectiveness rather than implementation. For example, the study by Wingard et al [22] reported the results of a pre-post study of the effectiveness of a phased, multicomponent intervention aimed at reducing hospital readmissions among patients at 26 US dialysis facilities (with patients from 18 nonrandomized facilities serving as controls). Readmissions were reduced after the introduction of the intervention (0.88-0.66 per patient-year) in the intervention facilities; however, the decline was not statistically significant compared with the decline in the control facilities (P=.26). Sparse information about implementation was provided, except that they found that only 42% of the patients were successfully contacted by assigned case managers within 30 days of discharge for the posthospitalization call or visit intervention component. Similarly, in another pre-post quality improvement study with a single-component intervention (postdischarge telephone follow-up) aimed at reducing readmissions in a dialysis population admitted to a single hospital, readmissions were reduced (from 28.4% to 24.6% in the 3-month project). However, in an audit, investigators found that only 71% of visits were assigned to a manager; of these, only 80% recorded a patient call attempt, and of these, 38% of patients completed the interview [23]. Although our intervention did not include a direct patient contact component, these prior results highlight the challenges for staff, particularly nurses, in dedicating time to such tasks, even when the tasks are supported by health care administration and assigned as part of the clinical workflow. This difficulty is exacerbated by inadequate dialysis nurse staffing levels, which have been shown to be predictive of hospital readmission [24,25].

Limitations

Limitations other than those noted previously deserve mention. First, the design of *DialysisConnect* precluded the inclusion of nephrologists who attended Emory Dialysis and EUHM as users in both settings; thus, they were only able to log in as dialysis facility users. Second, the user surveys had low response rates, which may have biased our results in either direction; however, we were able to collect data from users versus nonusers and from both settings. Third, some suggested changes to the system may have increased its use but were very labor intensive for our initial, limited pilot, including adding the ability for dialysis providers to *start* a hospitalization event when they send the patient to the hospital and incorporate information from the emergency department. Fourth, *DialysisConnect* was developed without patient or surrogate input and was unlikely to address the sociodemographic factors associated with hospital readmission in previous studies [26,27]. Fifth, system use logs are temporary and potentially helpful information, such as pages visited more frequently, SMS text messages versus email preferences for alerts, and access via mobile versus web browsers, which cannot be determined; for future studies, these and similar systems could be modified to create permanent logs of system use to better inform implementation. Sixth, the number of users trained was difficult to track in the virtual environment, and it was even more difficult to track the level of engagement with training; future studies could include brief surveys about the level of confidence in using the system to inform training efforts. Finally, the COVID-19 pandemic both delayed and shortened our pilot and altered our approach to rollout and training. Although few providers named lack of time because of additional duties related to COVID-19 in our surveys (none by the end of the pilot), it is likely that the pandemic also had some impact on provider behavior and willingness to use *DialysisConnect*, especially given that a surge of patients receiving dialysis and hospitalized at EUHM for COVID-19 and related complications occurred during our pilot (January 2021 to February 2021).

Lessons Learned for Future Studies

Future work using *DialysisConnect* or similar systems could leverage the most commonly reported motivator: the potential to save time over current processes. Testimonials from current users, who were generally satisfied and reported that *DialysisConnect* not only provided informed care but also saved time, may be helpful in convincing reluctant users to invest initial time to learn the system and incorporate it into work processes. These users, who are peers rather than members of the research team and are current and frequent users of the system, would be ideal system champions and essential for successful implementation. When targeting potential users and champions, investigators should consider that APPs (if available) may be more likely to engage with the system than physicians. Dialysis nurses, who are critical to the postdischarge processes, could be encouraged and potentially incentivized to learn and use the system. In addition, social workers should be engaged directly to ensure that they understand their essential role in care transitions, as well as the ability of the system to inform timely recognition of the need for services. The addition of content highly relevant to social workers (such as transportation needs for follow-up appointments or requests for referrals for mental health services) might also improve engagement. As much as possible, involving hospital or dialysis facility administration to encourage providers to use the system and facilitate the workflow changes required could increase provider engagement. EHR integration, if possible, is ideal; however, the inclusion of automated file feeds where available, the use of checkboxes and drop-down lists to limit typing, and the inclusion of only the most critical elements for admission and discharge considerably minimizes the time spent in the system, which is a message that can be prioritized by site champions.

Conclusions

DialysisConnect is a flexible and adaptable intervention for enhancing care coordination between dialysis facilities and hospitals that demonstrates the feasibility, usability, and sustainability. In addition, this pilot implementation study informs strategies to overcome multiple potential barriers to the use of *DialysisConnect* and similar interventions. Future work in this area should consider appropriate system champions, identification of key users and tasks to be incorporated into the workflow, and site-specific methods to encourage provider behavior changes.

Acknowledgments

The research reported in this publication was supported by the National Institute of Diabetes and Digestive and Kidney Diseases of the National Institutes of Health under award number R18DK118467. The content is the sole responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

The authors would like to thank the providers and staff at Emory Dialysis and Emory University Hospital Midtown who facilitated this pilot, especially CG, Linda Turberville-Trujillo, Michelle Young, Kathy Oliver, Ifeoma Imonugo, Tomiwa Ishmail, and Kristy Hamilton, and the staff at Apex Health Innovations for developing *DialysisConnect* (Richard Dacre, Amber Webster, Alice Jordan, Alex Turnbull, Jasper Kirby, and John Scott) and facilitating compliance (Charlie Bonar). The authors would also like to thank Christian Park and Grace Xu for their assistance with the focus group sessions.

Authors' Contributions

LCP, BGJ, CMO, and RM conceived the idea of *DialysisConnect*. LCP obtained funding for the project and wrote the first draft of the manuscript. AEV led the initial focus groups. RM led the technical team to develop *DialysisConnect*. JPL, TM, CMO, and KJ served as the site champions. CG was the first and most active user of *DialysisConnect* and served in a peer champion role, encouraging staff at both sites to use the system. LCP, CH, AEV, and AK analyzed and interpreted the system, survey, focus group, and electronic health record data, respectively. All authors have read and approved the final manuscript.

Conflicts of Interest

RM is the chief executive officer of Apex Health Innovations, which provided the technical expertise to build DialysisConnect. DialysisConnect was not, and is currently not, a commercially available product.

Multimedia Appendix 1
Focus group feedback.
[DOCX File , 24 KB - formative_v6i6e36052_app1.docx]

Multimedia Appendix 2
User testing feedback.
[DOCX File , 28 KB - formative_v6i6e36052_app2.docx]

Multimedia Appendix 3
Additional screenshots.
[DOCX File , 921 KB - formative_v6i6e36052_app3.docx]

Multimedia Appendix 4
Hospital user guide.
[PDF File (Adobe PDF File), 2901 KB - formative_v6i6e36052_app4.pdf]

Multimedia Appendix 5
Dialysis user guide.
[PDF File (Adobe PDF File), 3061 KB - formative_v6i6e36052_app5.pdf]

Multimedia Appendix 6
Provider survey.
[DOCX File , 27 KB - formative_v6i6e36052_app6.docx]

References

1. Doty MM, Tikkanen R, Shah A, Schneider EC. Primary care physicians' role in coordinating medical and health-related social needs in eleven countries. Health Aff (Millwood) 2020 Jan;39(1):115-123. [doi: 10.1377/hlthaff.2019.01088] [Medline: 31821045]
2. USRDS 2021 Annual Data Report: Atlas of Chronic Kidney Disease and End-Stage Renal Disease in the United States. United States Renal Data System. Bethesda, MD, USA: National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases; 2021. URL: https://adr.usrds.org/2021 [accessed 2021-12-15]

3. Coleman EA. Falling through the cracks: challenges and opportunities for improving transitional care for persons with continuous complex care needs. J Am Geriatr Soc 2003 Apr;51(4):549-555. [doi: 10.1046/j.1532-5415.2003.51185.x] [Medline: 12657078]
4. Mathew AT, Strippoli GF, Ruospo M, Fishbane S. Reducing hospital readmissions in patients with end-stage kidney disease. Kidney Int 2015 Dec;88(6):1250-1260 [FREE Full text] [doi: 10.1038/ki.2015.307] [Medline: 26466320]
5. Plantinga LC, King L, Patzer RE, Lea JP, Burkart JM, Hockenberry JM, et al. Early hospital readmission among hemodialysis patients in the United States is associated with subsequent mortality. Kidney Int 2017 Oct;92(4):934-941 [FREE Full text] [doi: 10.1016/j.kint.2017.03.025] [Medline: 28532710]
6. Doshi S, Wish JB. Strategies to reduce rehospitalization in patients with CKD and kidney failure. Clin J Am Soc Nephrol 2021 Feb 08;16(2):328-334 [FREE Full text] [doi: 10.2215/CJN.02300220] [Medline: 32660962]
7. Plantinga LC, Masud T, Lea JP, Burkart JM, O'Donnell CM, Jaar BG. Post-hospitalization dialysis facility processes of care and hospital readmissions among hemodialysis patients: a retrospective cohort study. BMC Nephrol 2018 Jul 31;19(1):186 [FREE Full text] [doi: 10.1186/s12882-018-0983-5] [Medline: 30064380]
8. Manley HJ, Aweh G, Weiner DE, Jiang H, Miskulin DC, Johnson D, et al. Multidisciplinary medication therapy management and hospital readmission in patients undergoing maintenance dialysis: a retrospective cohort study. Am J Kidney Dis 2020 Jul;76(1):13-21. [doi: 10.1053/j.ajkd.2019.12.002] [Medline: 32173107]
9. Erickson KF, Winkelmayer WC, Chertow GM, Bhattacharya J. Physician visits and 30-day hospital readmissions in patients receiving hemodialysis. J Am Soc Nephrol 2014 Sep;25(9):2079-2087 [FREE Full text] [doi: 10.1681/ASN.2013080879] [Medline: 24812168]
10. 2015 Annual Report. Georgia Health Information Network. 2016. URL: https://dch.georgia.gov/document/document/fy-2015-annual-report/download [accessed 2021-12-15]
11. Hersh WR, Totten AM, Eden KB, Devine B, Gorman P, Kassakian SZ, et al. Outcomes from health information exchange: systematic review and future research needs. JMIR Med Inform 2015 Dec 15;3(4):e39. [doi: 10.2196/medinform.5215] [Medline: 26678413]
12. 42 CFR part 412 (§412.150 through §412.154) – Hospital Readmissions Reduction Program. Centers for Medicare & Medicaid Services. 2012. URL: https://www.law.cornell.edu/cfr/text/42/412.154 [accessed 2021-11-01]
13. End-Stage Renal Disease Prospective Payment System, Quality Incentive Program, and Durable Medical Equipment, Prosthetics, Orthotics, and Supplies: Final Rule. Fed Register. 2014. URL: https://www.federalregister.gov/documents/2014/11/06/2014-26182/medicare-program-end-stage-renal-disease-prospective-payment-system-quality-incentive-program-and [accessed 2021-11-01]
14. Patient Protection and Affordable Care Act. 42 U.S.C. § 18001. 2010. URL: https://www.govinfo.gov/content/pkg/PLAW-111publ148/html/PLAW-111publ148.htm [accessed 2021-11-01]
15. Kelly YP, Kuperman GJ, Steele DJ, Mendu ML. Interoperability and patient electronic health record accessibility: opportunities to improve care delivery for dialysis patients. Am J Kidney Dis 2020 Sep;76(3):427-430. [doi: 10.1053/j.ajkd.2019.11.001] [Medline: 31973909]
16. Patzer RE, McPherson L, Redmond N, DuBay D, Zayas C, Hartmann E, et al. A culturally sensitive Web-based intervention to improve living donor kidney transplant among African Americans. Kidney Int Rep 2019 Sep;4(9):1285-1295 [FREE Full text] [doi: 10.1016/j.ekir.2019.05.771] [Medline: 31517147]
17. Vandenberg AE, Jaar BG, James KP, Lea J, O'Donnell C, Masud T, et al. Making sense of DialysisConnect: a qualitative analysis of stakeholder viewpoints on a Web-based information exchange platform to improve care transitions between dialysis clinics and hospitals. BMC Med Inform Decis Mak 2021 Feb 09;21(1):47 [FREE Full text] [doi: 10.1186/s12911-021-01415-y] [Medline: 33563290]
18. May C, Finch T. Implementing, embedding, and integrating practices: an outline of Normalization Process Theory. Sociology 2009 Jun 15;43(3):535-554. [doi: 10.1177/0038038509103208]
19. Finch TL, Rapley T, Girling M, Mair FS, Murray E, Treweek S, et al. Improving the normalization of complex interventions: measure development based on Normalization Process Theory (NoMAD): study protocol. Implement Sci 2013 Apr 11;8:43 [FREE Full text] [doi: 10.1186/1748-5908-8-43] [Medline: 23578304]
20. Grimshaw JM, Shirran L, Thomas R, Mowatt G, Fraser C, Bero L, et al. Changing provider behavior: an overview of systematic reviews of interventions. Med Care 2001 Aug;39(8 Suppl 2):II2-I45. [Medline: 11583120]
21. Gupta B, Li D, Dong P, Acri MC. From intention to action: a systematic literature review of provider behaviour change-focused interventions in physical health and behavioural health settings. J Eval Clin Pract 2021 Dec;27(6):1429-1445. [doi: 10.1111/jep.13547] [Medline: 33565177]
22. Wingard RL, McDougall K, Axley B, Howard A, O"Keefe C, Armistead N, et al. Right TraC™ post-hospitalization care transitions program to reduce readmissions for hemodialysis patients. Am J Nephrol 2017;45(6):532-539. [doi: 10.1159/000477325] [Medline: 28531888]
23. Briscoe GT, Heerschap A, Kane CF, Quatrara BD. Using post-discharge telephone follow-up by nephrology nurses to reduce 30-day readmissions and post-discharge complications for adult patients on hemodialysis. Nephrol Nurs J 2018;45(3):243-267. [Medline: 30304617]

24. Chen Y, Rhee C, Senturk D, Kurum E, Campos L, Li Y, et al. Association of US dialysis facility staffing with profiling of hospital-wide 30-day unplanned readmission. Kidney Dis (Basel) 2019 Jun;5(3):153-162 [FREE Full text] [doi: 10.1159/000496147] [Medline: 31259177]
25. Paulus A, Baernholdt M, Kear T, Jones T, Thacker L. Factors associated with hospital readmissions among U.S. dialysis facilities. J Healthc Qual 2022;44(2):59-68. [doi: 10.1097/JHQ.0000000000000300] [Medline: 34191751]
26. Adisa O, Jaar BG, Masud T, Sahlie A, Obadina C, Ang J, et al. Association of social worker-assessed psychosocial factors with 30-day hospital readmissions among hemodialysis patients. BMC Nephrol 2018 Dec 17;19(1):360 [FREE Full text] [doi: 10.1186/s12882-018-1162-4] [Medline: 30558578]
27. Hickson LJ, Thorsteinsdottir B, Ramar P, Reinalda MS, Crowson CS, Williams AW, et al. Hospital readmission among new dialysis patients associated with young age and poor functional status. Nephron 2018;139(1):1-12 [FREE Full text] [doi: 10.1159/000485985] [Medline: 29402792]

Abbreviations

APP: advanced practice provider
EHR: electronic health record
EUHM: Emory University Hospital Midtown

Edited by A Mavragani; submitted 29.12.21; peer-reviewed by J Wacksman, J Knitza, L Golestaneh, D Hynes; comments to author 24.04.22; revised version received 02.05.22; accepted 19.05.22; published 10.06.22.

Please cite as:
Plantinga LC, Hoge C, Vandenberg AE, James K, Masud T, Khakharia A, Gray C, Jaar BG, Lea JP, O'Donnell CM, Mutell R
A Web-Based, Provider-Driven Mobile App to Enhance Patient Care Coordination Between Dialysis Facilities and Hospitals: Development and Pilot Implementation Study
JMIR Form Res 2022;6(6):e36052
URL: https://formative.jmir.org/2022/6/e36052
doi:10.2196/36052
PMID:35687405

Original Paper

Time-Varying Associations Between Device-Based and Ecological Momentary Assessment–Reported Sedentary Behaviors and the Concurrent Affective States Among Adolescents: Proof-of-Concept Study

Jennifer Zink[1], PhD; Chih-Hsiang Yang[2], PhD; Jasmin M Alves[3], PhD; Kelsey L McAlister[1], MS; Jimi Huh[1], PhD; Mary Ann Pentz[1], PhD; Kathleen A Page[3], MD; Genevieve F Dunton[1,4], MPH, PhD; Britni R Belcher[1], MPH, PhD

[1]Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles, CA, United States
[2]Department of Exercise Science, Arnold School of Public Health, University of South Carolina, Columbia, SC, United States
[3]Department of Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, United States
[4]Department of Psychology, University of Southern California, Los Angeles, CA, United States

Corresponding Author:
Britni R Belcher, MPH, PhD
Department of Population and Public Health Sciences
Keck School of Medicine
University of Southern California
2001 N Soto St
Los Angeles, CA, 90032
United States
Phone: 1 323 442 8225
Email: bbelcher@usc.edu

Abstract

Background: Previous studies on affective state–sedentary behavior (SB) associations have not accounted for their potentially time-varying nature and have used inconsistent SB measurement modalities. We investigated whether the strength of the associations between affective states and SB varied as a function of the time of day and by SB measurement modality (device-measured SB vs ecological momentary assessment–reported screen-based SB) in youth.

Objective: This study aimed to establish a proof of concept that SB–affective state associations may not be static during the day. In addition, we aimed to inform the methodology of future work, which may need to model associations as functions of the time of day and carefully consider how SB is operationalized or measured.

Methods: A total of 15 adolescents (age: mean 13.07, SD 1.03 years; 10/15, 67% female; 6/15, 40% Hispanic; 10/15, 67% healthy weight) wore thigh-mounted activPAL accelerometers and simultaneously reported their screen-based SBs and concurrent positive and negative affective states via ecological momentary assessment for 7 to 14 days (N=636 occasions). Time-varying effect models (varying slopes) examined how each measure of SB was associated with concurrent affective states from 7 AM to 8 PM.

Results: Time-varying effect model plots revealed that these associations varied in strength throughout the day. Specifically, device-based SB was related to greater concurrent negative affect only after approximately 5 PM and was unrelated to concurrent positive affect. Screen-based SB was related to greater concurrent negative affect only from 7 AM to approximately 9 AM. This was also related to greater concurrent positive affect from 7 AM to approximately 9:30 AM and from approximately 3 PM to approximately 7 PM.

Conclusions: We provide preliminary evidence to suggest that future confirmatory studies investigating the SB–affective state relationship should consider the time-varying nature of these associations and SB measurement modality. There may be critical time windows when specific types of SBs co-occur with affect, suggesting that interventions may need tailoring to the time of day and type of SB if future studies using similar methodologies can replicate our findings.

(JMIR Form Res 2022;6(6):e37743) doi:10.2196/37743

KEYWORDS

accelerometry; intensive longitudinal data; mood; youth; mobile phone

Introduction

Background

Sedentary behaviors (SBs), such as screen-based behaviors, are highly prevalent among adolescents and are associated with poor health outcomes [1]. Over 30% and 40% of adolescents in the United States report ≥3 hours of daily television viewing and computer use, respectively [2]. The pervasiveness of SB among young people may be attributed to factors such as urbanization [3], technological advancements [4], and the development of social media [5]. Therefore, opportunities for being sedentary will continue to surround youth. Interventions aimed at reducing SB have thus far been relatively ineffective [6]. Understanding the potential correlates of SBs as they occur naturally in everyday life may be important for the development of effective intervention strategies aimed at reducing sedentary time.

Affective states may be an important factor to consider in future behavior change interventions. Acutely, positive affective states can be related to salubrious behaviors such as physical activity [7-9], whereas negative affective states are associated with unfavorable behaviors such as fast-food consumption [10] and cigarette smoking [11] among youth. Therefore, it is plausible that affective states and SB may co-occur during adolescence, a developmental period when sedentariness increases, and affect can become less positive and more variable [12,13]. However, extant studies on the association between affective states and SB among youth have yielded inconsistent findings [8,14,15].

Inconsistencies in the literature thus far may be attributed to unaccounted-for complexities in the association between affective states and SBs, such as their potentially time-varying nature. For example, evidence suggests that SB specifically in the evening may be linked to emotional health outcomes, including worse affect [16,17]. Ecological momentary assessment (EMA) is a data collection method whereby participants report their affective states and behaviors via mobile devices as they occur naturally in real time. It is an ideal method for capturing acute changes in affective states, SBs, and their associations with one another across the day [18]. The time-varying effect model (TVEM) is also being increasingly used in the field of behavioral medicine, especially for leveraging EMA data [19,20]. The TVEM optimizes the repeated-measures data structure of EMA to model dynamic associations over time (across the day). This allows investigators to pinpoint the specific time windows in which associations may be strongest or weakest between 2 variables [21,22]. To our knowledge, previous investigations of the within-day associations between affective states and SB have yet to simultaneously use EMA and the TVEM to assess whether the strength of these associations varies across the chronological time of day. Therefore, our understanding of the potentially time-specific co-occurrence of affective states and SB is limited.

We also have a limited understanding of the potential associations at hand because previous studies have used inconsistent measures or operationalizations of SB. Studies of device-based SB (capturing time spent sitting or lying down) and affective states among youth indicate that negative affect is directly related to SB, whereas positive affect is inversely related to SB [8,14]. In contrast, another study on the associations between EMA-reported screen-based SB (capturing behaviors done while sitting or lying) and affective states among youth yielded null findings [15]. It is unclear whether these conflicting findings are because of measurement and recall errors associated with subjective measures of behavior or whether device-based and subjective SB represent distinct constructs that are differentially related to affective states. Thus, studies with more rigorous approaches, such as those that combine device-based and EMA-reported SB, are needed to elucidate the possible associations at hand.

Objectives

Taken together, the next step in this area of research is to examine the potential differences in the strength of the within-day associations between affective states and SB by the time of day and operationalization of SB. Therefore, our goal was to provide a proof-of-concept study that would inform the methodology of future work in the topic area of affective states and SB, which may need to account for the time-varying nature of these associations and carefully consider how SB is operationalized moving forward. On the basis of the abovementioned work [8,14,16,17], we hypothesized that higher negative affect, lower positive affect, and SB would co-occur in the evening hours, specifically for device-based SB. The findings of this study can increase our understanding of the relationships at hand by (1) identifying the specific time points within the day that these associations are most likely to occur and (2) identifying which operationalization or measure of SB relates to affective states the most strongly.

Methods

Participants and Recruitment

The participants (N=15) in this study were a subset of participants from the Mothers' and Their Children's Health (MATCH) cohort study of maternal stress and their children's obesity risk, which took place in the broader Los Angeles, California, United States, metropolitan area [23]. MATCH study participants were recruited via flyers and in-person research staff visits to public elementary schools and community events. The inclusion criteria for mother-child dyads in the MATCH study were as follows: (1) the child is in third to sixth grade at baseline, (2) more than half of the child's custody belongs to the mother, and (3) both mother and child can read English or Spanish. Dyads were excluded from the MATCH cohort if the mother or the child (1) was taking medications for thyroid function or psychological conditions, (2) had a health condition that limited physical activity, (3) was enrolled in a special education program, (4) was currently using oral or inhalant corticosteroids for asthma, (5) was pregnant, (6) was classified as underweight by a BMI percentile of <5% adjusted for sex

and age (child only), or (7) worked >2 weekday evenings (between 5 and 9 PM) per week or >8 hours on any weekend day (mother only). The detailed MATCH study protocol can be found elsewhere [23].

Participants enrolled in the MATCH study were recruited for the Sedentary Behavior and Health Outcomes Study, an in-laboratory randomized crossover trial investigating the metabolic effects of interrupting sitting (ClinicalTrials.gov NCT03153930). To be eligible for the Sedentary Behavior and Health Outcomes Study (and therefore this substudy), participants were required to be enrolled in the MATCH cohort across all 6 waves (3 years) and be in good general health. Youth with cardiac or pulmonary disease, allergies to metals, evidence of type 2 diabetes, and endocrinologic disorders leading to obesity or taking medications for attention-deficit/hyperactivity disorder were excluded.

Procedures

Data collection for this study was conducted from May to November 2018. The work presented here is a secondary analysis of pilot data with the primary goal of demonstrating the scalability of combining EMA and continuous glucose monitoring [24]. Enrollment for this pilot study was performed on a rolling basis until the target sample size (N=15) was met. At the baseline screening visit, the participants and their parents provided assent and consent, respectively. Participants reported demographic characteristics, including age, sex, ethnicity, and highest maternal education achieved, which was used as a proxy for socioeconomic status. Anthropometric measurements (height in centimeters and weight in kilograms) were collected in duplicate by trained study staff and used to calculate the age-adjusted BMI percentile using the Centers for Disease Control and Prevention *EpiInfo* tool.

Eligible participants were then instructed to wear an activPAL micro4 accelerometer (PAL Technologies) on their right thigh for 24 hours per day for the next 7 complete days. They were instructed to wear the device at all times, including on school and summer camp days, during extracurricular activities, and during showering or bathing. The study staff placed a waterproof cover on the activPAL and taped the devices to the mid–right thigh to ensure proper placement and minimize participant removal of the device during the assessment period. ActivPALs have been validated for use with youth, and as they are thigh mounted, they can differentiate between sitting and standing, making them ideal for capturing SB [25].

Participants were also provided a Moto G mobile phone (Motorola Mobility) with the *Movisens* EMA app predownloaded for the duration of the study. On weekend days, each participant received 7 random EMA prompts during specified 1-hour time windows between 7 AM and 8 PM. On weekdays, participants received 4 random prompts during specified 1-hour time windows between 7 AM and 8 PM (except during school or summer camp hours, defined as weekdays from 8 AM to 3 PM). Therefore, participants received up to 34 EMA prompts (n=14, 41% on weekend days and n=20, 59% on weekdays) across the 7-day assessment period. The 1-hour EMA prompt time windows are listed in Multimedia Appendix 1. The EMA prompts occurred during the same 7 days when the participants wore the activPAL device. The EMA mobile device chimed and vibrated to prompt the participant to stop their current activity and answer the EMA survey, which took approximately 2 minutes to complete. If a participant did not respond to the EMA survey after the initial alert, there were up to 5 reminder signals within 20 minutes of the initial alert. The EMA survey expired after the fifth alert was ignored. At each prompt, participants were asked to report on their concurrent SBs and affective states. All EMA survey responses were date- and time-stamped.

As part of the Sedentary Behavior and Health Outcomes Study, after the initial 7-day assessment period described previously, participants were asked to return the activPALs and mobile devices to the study team and complete the same 7-day study protocol again (after a washout period ranging from 1 week to approximately 1 month). Therefore, all 15 participants had the opportunity to contribute 14 assessment days (2 separate 7-day observational periods) to the data in this study. Of the 15 participants, 14 (93%) contributed 14 assessment days of data and 1 (7%) participant contributed 7 assessment days of data, all of which were included in the present analyses. Depending on the randomization order from the Sedentary Behavior and Health Outcomes Study in-laboratory randomized trial, all participants were additionally given a wrist-worn activity monitor (LYCOS Life). This device was programmed to prompt participants to interrupt their SB (eg, with walking) every 30 minutes using vibrations for either the first or second assessment week. During the assessment periods when the wrist-worn activity monitor was not assigned to the participants, the study staff instructed the participants to proceed with their normal daily routines. Of the 15 participants, 7 (47%) received LYCOS Life during the first assessment week and 8 (53%) received LYCOS Life during the second observational week. Randomization order (whether the participant received LYCOS Life during the first or second assessment week) did not differ by participant characteristics (age: P=.80; sex: P=.71; highest maternal education: P=.18; ethnicity: P=.83; weight status: P=.71).

Measures

Device-Based SB

All sleep and nonwear times were removed before analyzing activPAL-measured SB. Nonwear was defined as ≥60 consecutive minutes of 0 counts [26], whereas valid days were defined as ≥10 hours of valid wear time during waking hours [27]. SB was defined as activities requiring ≤1.5 metabolic equivalents [28], and the total number of minutes spent in SB was calculated for the matched 15-minute time window before the EMA prompt. This 15-minute time frame is consistent with previous studies of affective states and SB among youth [19,29] and was considered the smallest meaningful amount of time that matched the presented EMA screen-based SB item wording (eg, *just before the phone went off*).

Self-reported Screen-Based SB

Via EMA, participants were asked to select the primary SB that they were currently engaged in at the time of the EMA prompt (*just before the phone went off*). The response options were

television, movies, or videos; social media (Facebook, Snapchat, Instagram, and Tumblr); videogames; computer or tablet use; homework or reading; hanging out or chatting; art, painting, or coloring; riding in the car or bus; and none of these items. This item was dummy coded into 1 dichotomous screen-based SB variable, where 1=*yes* to screen-based SB (television, movies, or videos; social media; videogames; and computer or tablet use) versus 0=*no* to screen-based SB (homework/reading; hanging out or chatting; art, painting, or coloring; riding in the car or bus; and none of these things).

Affective States

EMA questions prompted participants to report on their current affective states (*just before the phone went off*) based on five items of the Positive and Negative Affect Schedule–Child short form—stressed, mad, sad, happy, and joyful—consistent with previous EMA studies of affective states among youth [10,19]. The response options ranged from 0 to 3 (0=*not at all*, 1=*a little*, 2=*quite a bit*, and 3=*extremely*). The responses for stressed, mad, and sad (3 items) were averaged to create a continuous score for negative affect (within-subject internal consistency reliability, ω=0.81), and the responses for happy and joyful (2 items) were averaged to create a continuous score for positive affect (ω=0.90). Therefore, negative and positive affect could each range from 0 to 3 at any given EMA prompt, with higher scores indicating higher negative or positive affect.

Covariates

Time-invariant covariates were selected a priori based on previous work showing associations between SB and symptoms of emotional disorders, including age (continuous; years), sex (dichotomous; female, 1=yes vs 0=no), ethnicity (dichotomous; Hispanic, 1=yes vs 0=no), socioeconomic status (dichotomous; maternal education college or higher, 1=yes vs 0=no), and weight status (based on BMI percentile; dichotomous; overweight/obese, 1=yes vs 0=no) [30-32]. In addition, the time-varying covariate, day of the week (dichotomous; weekend, 1=yes vs 0=no), was included in all models. EMA-reported physical activity (dichotomous; any physical activity, 1=yes vs 0=no), environmental context (dichotomous; indoors, 1=yes vs 0=no), social context (dichotomous; alone, 1=yes vs 0=no), and experimental condition (LYCOS Life Band, 1=yes vs 0=no) were each tested one at a time as covariates and were retained in the models if they were significantly associated with the outcome at the $P<.05$ level. Multimedia Appendix 2 provides a more detailed description of each EMA item.

Statistical Analysis

Frequencies or means were calculated for participant characteristics, affective states, activPAL-measured SB (in the 15-minute time window before the EMA prompt), and EMA-reported screen-based SB (*yes* to any screen-based SB *just before the phone went off*). Cross-tabulations were used to calculate the mean affective state score based on yes or no reports of screen-based SB. Sample- and individual-level EMA prompt compliances were calculated as the proportion of prompts completed out of the total number of prompts sent to the participants. To better understand the potential patterns of data missingness, separate multilevel logistic regression models tested whether participant age, sex, ethnicity, weight status (healthy weight vs overweight or obese), maternal education, day of the week, week-level positive affect, and week-level negative affected predicted momentary EMA prompt compliance (prompt completed; yes vs no) and valid activPAL wear (valid day; yes vs no). Multimedia Appendix 3 presents the descriptive statistics of EMA prompt compliance.

TVEMs, which are uniquely suited for the analysis of intensive, repeated measures (eg, time-stamped activPAL and EMA data), were used to model the associations between affective states and SB across the day. TVEMs are designed to test changes in the strength of the association between the predictor and outcome over time, which is modeled nonparametrically. Moreover, TVEMs can accommodate an unequal temporal spacing of observations and an unequal number of observations per participant because of missing observations, which is common in EMA studies [22]. The TVEM results are presented graphically, where time is on the x-axis. From a single TVEM, 2 graphical results were produced: (1) an intercept function, which represents momentary levels of affect (the outcome) for participants with average levels of all covariates at a given time, and (2) a slope function, which represents the adjusted estimate of the association (β) between SB and concurrent affect at any given time. In this study, the solid line in the figures representing the graphical results represents the point estimate, whereas the dashed lines represent the corresponding 95% CIs. A CI (dashed lines) that does not overlap with zero at any moment in time for the slope function indicates a significant association between the predictor (SB) and outcome (affective states) during a specific time interval. All models in this study presented results from 7 AM to 8 PM because of the EMA sampling protocol.

TVEMs were conducted using %TVEM SAS macro [21]. The TVEMs were fitted using the default setting that applied the penalized truncated power spline (P-spline) technique with 10 knots (dividing points) for computational flexibility and efficiency [19]. In contrast to the unpenalized (B-spline) technique, the P-spline method uses an automated model selection procedure, making it the preferred method for fitting TVEMs that can have complex coefficient functions [21,22]. Therefore, the P-spline approach is more appropriate for modeling momentary within-day changes captured in EMA studies on health behavior [33].

Two empty TVEMs (with only an intercept function and an error term as predictors) were used to describe the unadjusted average levels of (1) negative affect and (2) positive affect reported across the day. To model the associations of interest, four conditional TVEMs were used to assess the changes in the strength of the association between (1) activPAL-measured SB and concurrent negative affect across the day, (2) EMA-reported screen-based SB and concurrent negative affect across the day, (3) activPAL-measured SB and concurrent positive affect across the day, and (4) EMA-reported screen-based SB and concurrent positive affect across the day.

As one of our exposure variables of interest (screen-based SB) encompassed smartphone use, sensitivity analyses were conducted by removing observations from participants (2/15, 13%) who had no prior smartphone ownership. This was

performed to assess how providing a study smartphone to those who may otherwise not have had regular access to a smartphone could have influenced our study findings. To understand how the day of the week may have influenced our results, additional sensitivity analyses were conducted to examine the acute associations on weekend days only. Furthermore, as evidence suggests that engagement in screen-based SB may differ on weekend days versus weekdays among youth [34], we calculated frequencies of the reported screen-based SB stratified by the time of day and day of the week (weekend day vs weekday) to investigate whether this was observed in our sample. All analyses were conducted using SAS (version 9.4).

Ethics Approval

All study procedures were approved by the University of Southern California Institutional Review Board (HS-17-00126).

Results

ActivPAL and EMA Compliance

Table 1 shows the characteristics of the study sample. Of the 1030 EMA prompts received, participants completed 636 (61.74% sample-level compliance); thus, the analytic sample size was 636 for TVEMs that leveraged only EMA items. Participant-level EMA compliance ranged from 32.47% to 88%. The multilevel logistic regression analyses of momentary prompt-level EMA compliance indicated that participant characteristics (age, sex, ethnicity, maternal education, weight status, person mean negative affect, and person mean positive affect) were unrelated to EMA prompt compliance (all $P>.18$). The day of the week was also unrelated to momentary EMA prompt compliance ($P=.39$). Participants were more likely to complete EMA prompts later in the day (odds ratio 1.08, 95% CI 1.01-1.14; $P=.02$).

Of the 636 completed EMA prompts, 94 (14.8%) were removed as they were paired with activPAL observations that occurred on nonvalid days. This yielded an analytic sample size of 542 activPAL-matched EMA prompts for TVEMs, where device-based SB (in the past 15 minutes) was the predictor of affective states. The multilevel logistic regression analyses of valid activPAL wear at the day level indicated that participant characteristics, including person mean affective states, were unrelated to valid wear time (all $P>.15$). The day of the week was related to valid wear time, with valid days being less likely to occur on weekend days (odds ratio 0.25, 95% CI 0.17-0.37; $P<.001$).

Table 1. Characteristics of the study sample and descriptive statistics of main study variables (participants: N=15; 636 ecological momentary assessment prompts).

Characteristics	Values
Age (years), mean (SD)	13.07 (1.03)
Sex (female), n (%)	10 (67)
Ethnicity (Hispanic), n (%)	6 (40)
Highest maternal education (college and above), n (%)	11 (73)
Weight status (healthy weight), n (%)	10 (67)
BMI percentile, mean (SD)	55.42 (32.05)
ActivPAL SB[a,b] (minutes), mean (SD)	11.88 (4.24)
Screen-based SB (yes), n (%)	257 (40.4)
Negative affect, mean (SD)	0.29 (0.60)
Positive affect, mean (SD)	1.58 (0.97)

[a]ActivPAL-measured SB in the 15-minute window before the ecological momentary assessment prompt.

[b]SB: sedentary behavior.

Descriptive Statistics

Table 1 presents descriptive statistics of the main study variables. Across all answered EMA prompts, participants reported an average negative affect of 0.29 (SD 0.60) and an average positive affect of 1.58 (SD 0.97). On occasions when screen-based SBs were reported, the mean negative affect was 0.40 (SD 0.76), whereas it was 0.22 (SD 0.44) on occasions when screen-based SBs were not reported. On occasions when screen-based SBs were reported, the mean positive affect was 1.75 (SD 0.98), whereas it was 1.47 (SD 0.94) on occasions when screen-based SBs were not reported.

SB and Negative Affect

The intercept-only (unadjusted for covariates) TVEM plot for negative affect is presented in Figure 1 (panel A). The mean level of EMA-reported negative affect remained steady at around 0.30 across the daily EMA-prompting period (7 AM to 8 PM). The highest levels of negative affect were reported just before 10 AM (mean negative affect 0.35, 95% CI 0.12-0.58) and the lowest levels of negative affect were reported at approximately 5 PM (mean negative affect 0.25, 95% CI 0.09-0.41). The intercept functions in Figure 2 (panel A and panel C) indicate that negative affect was 0.06 to 0.81 across the entire day, adjusting for the average levels of all covariates. The slope function in panel B (Figure 2) presents the time-varying acute associations between activPAL-measured SB in the 15-minute

window before the EMA prompt and EMA-reported negative affect across the day. SB was unrelated to negative affect until just after 5 PM when SB was directly related to concurrent negative affect until 8 PM (β range .01-.06) after adjusting for a priori covariates and environmental context. Panel D (Figure 2) presents the time-varying acute associations between EMA-reported screen-based SB and concurrent negative affect from 7 AM to 8 PM, indicating that there was a significant direct association from 7 AM to just before 9 AM (β range .29-.41) after adjusting for a priori covariates, social context, and environmental context. Physical activity and LYCOS Life were not significantly associated with the outcome and were therefore not retained in these models.

Figure 1. Intercept-only time-varying effect model plots depicting unadjusted average negative affect (panel A) and unadjusted average positive affect (panel B) from 7 AM to 8 PM (N=636). The solid red line represents the point estimate; the dashed gray lines represent the corresponding 95% CI.

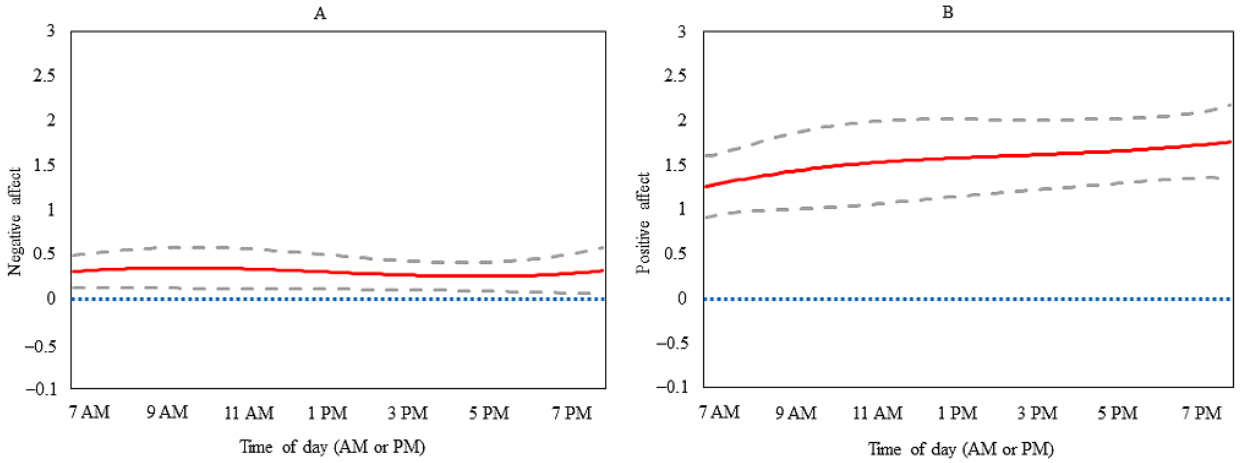
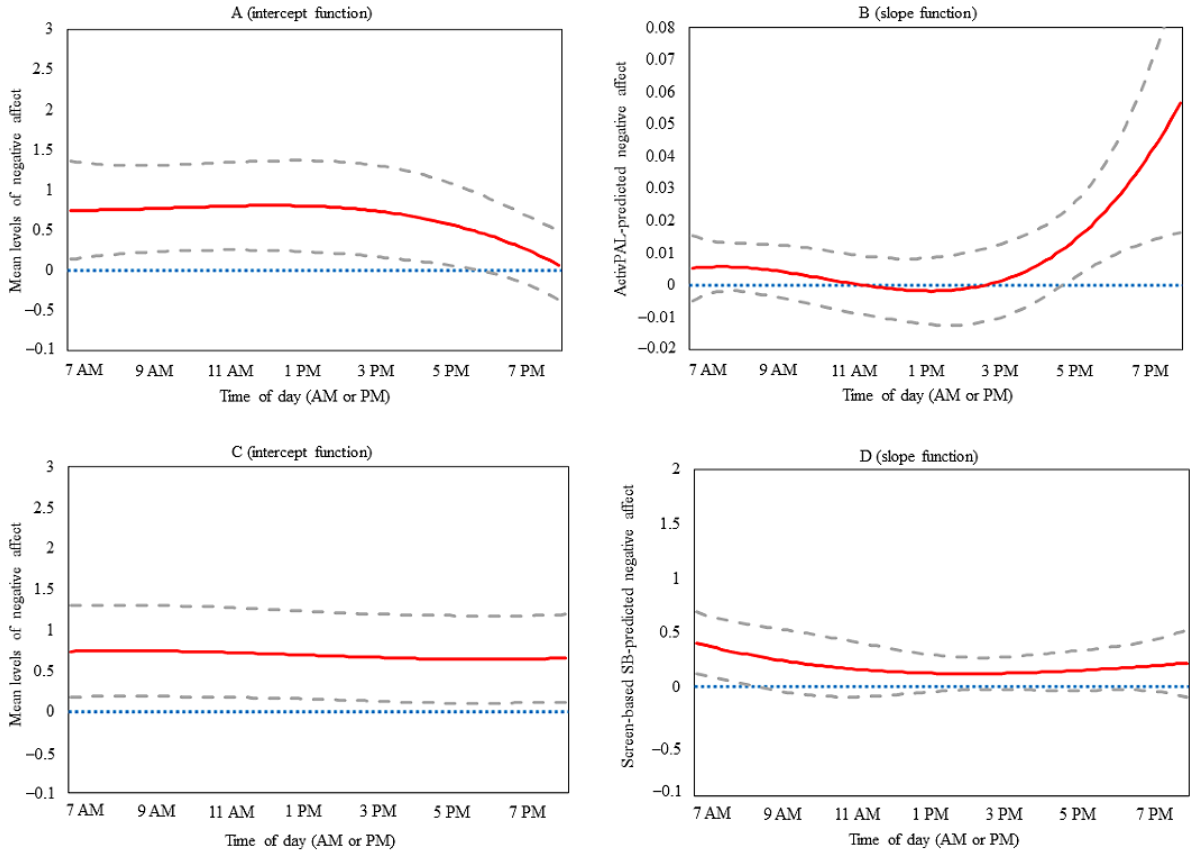


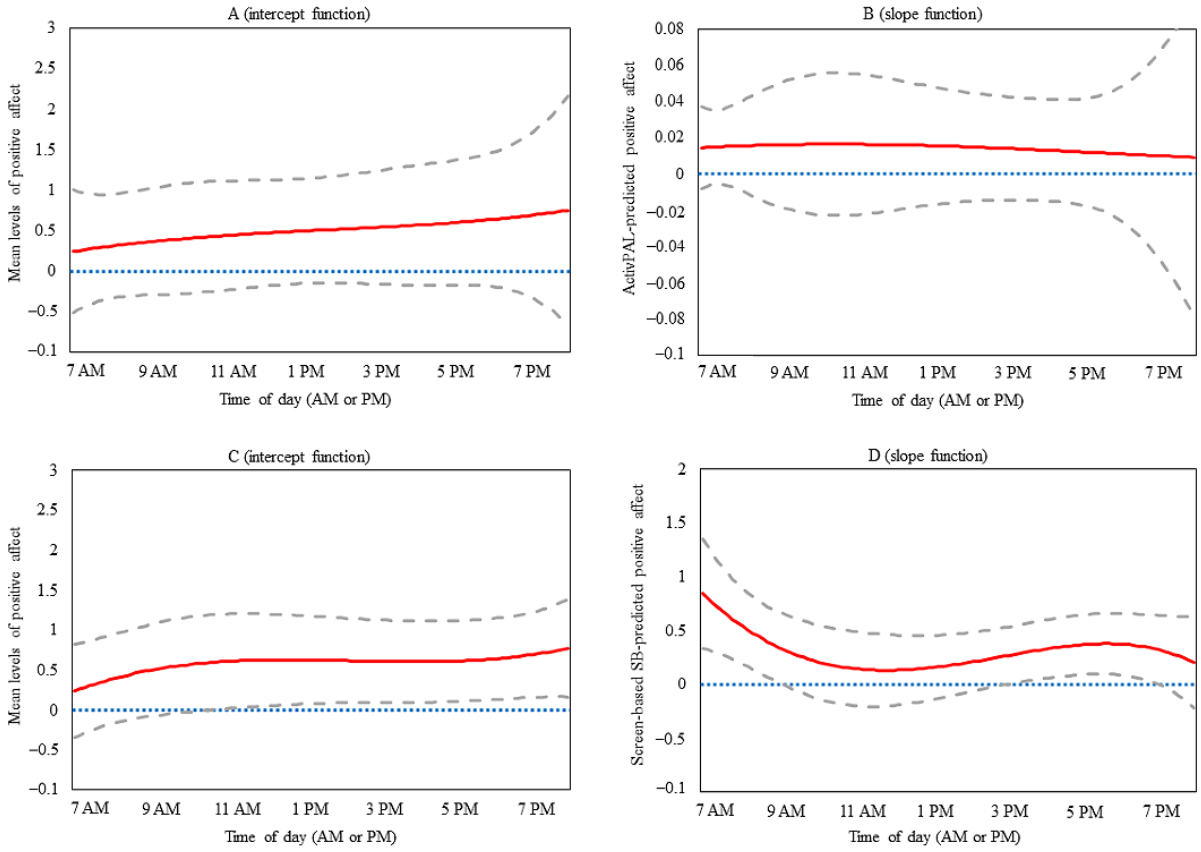
Figure 2. Time-varying effect model plots depicting the intercept and slope functions of the association between SB and concurrent negative affect from 7 AM to 8 PM. The intercept function represents momentary levels of negative affect, adjusted for covariates. The slope functions represent the adjusted estimate of the association (β) between SB and concurrent negative affect. Panels A and B present estimates from the activPAL model (N=542). Panels C and D present estimates from the ecological momentary assessment–reported screen-based SB model (N=636). The solid red line represents the point estimate; the dashed gray lines represent the corresponding 95% CI. SB: sedentary behavior.



SB and Positive Affect

The intercept-only (unadjusted for covariates) TVEM plot for positive affect is presented in Figure 1 (panel B). The mean level of EMA-reported positive affect slightly increased across the day from 7 AM (mean positive affect 1.25, 95% CI 0.90-1.59) to 8 PM (mean positive affect 1.76, 95% CI 1.35-2.18). The intercept functions in Figure 3 (panel A and panel C) indicate that positive affect was 0.22 to 0.78 across the entire day, adjusting for average levels of all covariates. The slope function (panel B, Figure 3) presents the time-varying acute associations between activPAL-measured sedentary time and positive affect, demonstrating that these associations were nonsignificant across the day (from 7 AM to 8 PM). Panel D (Figure 3) presents the time-varying acute associations between EMA-reported screen-based SB and positive affect across the day; significant direct associations were observed from 7 AM to just after 9 AM (β range .35-.88) and from just after 3 PM to just after 7 PM (β range .27-.38). Each of the abovementioned models was adjusted for a priori covariates but was not adjusted for additional potential covariates (eg, physical activity, environmental context, social context, and LYCOS Life) as they were not associated with the outcome.

Figure 3. Time-varying effect model plots depicting the intercept and slope functions of the association between SB and concurrent positive affect from 7 AM to 8 PM. The intercept function represents momentary levels of positive affect, adjusted for covariates. The slope functions represent the adjusted estimate of the association (β) between SB and concurrent positive affect. Panels A and B present estimates from the activPAL model (N=542). Panels C and D present estimates from the ecological momentary assessment–reported screen-based SB model (N=636). The solid red line represents the point estimate; the dashed gray lines represent the corresponding 95% CI. SB: sedentary behavior.



Sensitivity Analyses

After removing participants with no prior smartphone ownership (2/15, 13% of participants who collectively contributed 95/636, 14.9%, completed EMA prompts), the time-varying associations between activPAL-measured SB, screen-based SB, and concurrent affective states remained comparable with those presented previously. Therefore, providing a smartphone to those who otherwise might not have had regular access to a smartphone did not appear to influence our results, and these participants were retained in the final models. Multimedia Appendix 4 presents the frequency of EMA-reported screen-based SB by the time of day (EMA-prompting window) and day of the week (weekend day vs weekday). The frequency of screen-based SB did not differ by day of the week during any time of day (all chi-square $P>.05$). In addition, all models were rerun using data from weekend days only, and the results remained comparable with the main study findings. Therefore, all models presented data from weekdays and weekend days combined, with the day of the week as a covariate.

Discussion

Principal Findings

In this proof-of-concept study, we provided initial evidence to suggest that the associations between SB and concurrently reported affective states may differ by time of day and SB measurement modality. We did so by leveraging the TVEM, device-based activity monitoring, and EMA. We found that device-based SB was associated with more concurrent negative affect, possibly only in the evening. Our findings also indicated that device-based SB may be unrelated to concurrent positive affect across the entire day. Alternatively, EMA-reported screen-based SBs were related to more concurrent negative affect, possibly only in the morning. EMA-reported screen-based SBs also appeared to be related to more concurrent positive affect in the morning and late afternoon. Taken together, our results indicate that there may be critical windows during the day in which specific types of SBs tend to co-occur with affective states. This could have important intervention implications if future confirmatory studies using similar methodologies can replicate these findings.

Comparison With Prior Work

We demonstrated that the direct association between screen-based SBs and concurrent negative affect may only be

significant during morning hours (7 AM to approximately 9 AM). These findings are in line with a previous longitudinal study of adolescents, which found that screen-based SBs were bidirectionally associated with negative affective depressive symptoms but not with other types of depressive symptoms [35]. In contrast, although previous evidence suggests that engagement in screen-based SB, specifically in the evening, is related to negative affect, we did not observe such associations in the evening hours [17]. It is believed that engagement in screen-based SBs later in the day may influence sleep duration or quality, which in turn affects other factors such as mood and executive function [36]. However, our findings may not be consistent with this notion as the most commonly reported screen-based SB in our sample was television viewing (approximately 50% of all EMA prompts when engagement in screen-based SB was reported). Prior cross-sectional and longitudinal research suggests that compared with other forms of screen-based SB, television viewing is not as strongly related to emotional outcomes (and vice versa), perhaps because of its passive nature [37]. In contrast, nighttime engagement in other types of active screen-based SBs, such as computers and mobile phones, may be more strongly related to emotional health than television viewing [17,38,39]. Therefore, further research is needed on the potential for device- and time-specific associations between screen-based SB and negative affect, and the possible roles of sleep duration and quality should be further explored.

We also demonstrated that engagement in screen-based SBs may be related to more concurrent positive affect during the morning and afternoon hours. Depending on the time of day, adolescents may view screen-based SBs as pleasurable activities for coping with stressors [40,41]. Participants in our sample may have engaged in screen-based SBs in the morning hours to relax in preparation for the upcoming day at school (weekdays) or the day of structured organized activities (weekend days). Similarly, our sample may have engaged in screen-based SBs in the afternoon hours to attempt to alleviate stress from academics on weekdays and unwind from overscheduling on weekend days, which can be common during adolescence [42-44]. Longitudinal and bidirectional evidence across 1 year indicates that screen-based SBs are unrelated to positive affect in both directions [35]. Youth may choose behavioral coping strategies (ie, screen-based SBs) to manage or improve mood following a stressor; however, these coping strategies are considered maladaptive because while affective states may acutely improve, emotional and physical health are more likely to worsen long-term [45]. Therefore, the within-day association between engagement in screen-based SBs and higher positive affect may be transitory and likely does not accumulate into longer-term emotional or physical health benefits.

Our analyses of device-based SB in relation to affective states yielded differential findings compared with our analyses of EMA-reported screen-based SBs. Device-based SB was directly related to concurrent negative affect in the evening hours (approximately 5-8 PM) in this study. A previous free-living study of the acute (eg, past 30 minutes) associations between device-based sedentary time and negative affect among youth did not find that within-person deviations from one's usual sedentary time were related to negative affect [14]. Our use of the TVEM provides insight into a potential source of this inconsistency in the study findings. Other modeling methods typically used for multilevel data (EMA data) are parametric and may impose linearity on conceptual time, whereas associations across time may be nonlinear [46,47]. Therefore, when the strength of the association of interest may nonparametrically differ as a function of time, TVEMs may reveal associations that other common parametric modeling methods may not be able to capture [48]. Given our findings, the acute associations between SB and affective states appear to be nonlinear and nonparametric functions of the time of day. Therefore, future studies should consider taking a TVEM approach to understand the within-day associations between behaviors and affective states. Additional evidence from studies such as ours is needed to identify possible time windows of opportunity (when associations between behavior and affect are strongest) for intervention strategies to be delivered.

The differential findings between screen-based SB and device-based sedentary time also highlight that each is a distinct, yet interrelated, nuance of a larger behavior, broadly referred to as SB. A previous study among youth found that EMA-reported screen-based SBs were highly correlated with device-based sedentary time [49], and ancillary analyses of our sample also support this finding. Together, this suggests that our participants minimally misreported their engagement in screen-based SB via the EMA surveys. Therefore, the differences in associations by SB measurement modality (eg, EMA-reported screen-based SBs vs device-based sedentary time) observed in our study were likely not because of recall biases or errors. Rather, our study supports the notion that the behaviors performed while sedentary may uniquely relate to affective states in addition to objective time spent sitting. Altogether, our findings suggest that future investigations of SB–affective state associations should consider approaches that combine device-based measures of sedentary time and self-reported engagement in screen-based SBs, as they provide complementary behavioral information.

Strengths and Limitations

The strengths of this study include the real-time, repeated-measures data collection methods that were leveraged, which allowed us to model complex temporal associations and within-day changes in the association between SB and affective states using the TVEM. As EMA and accelerometry are data collection strategies that capture data in a naturalistic setting, this study was ecologically valid. Furthermore, to the best of our knowledge, this was the first study that combined EMA and device-based SB to directly examine how the operationalization of SB may have influenced the strength of the associations at hand.

However, there are also limitations that warrant further discussion. For brevity of our EMA surveys, we only used 5 items from the Positive and Negative Affect Schedule–Child survey to capture affective states. Although this is consistent with previous EMA studies on youth [10,19], future work could consider using more survey items to capture affective experiences. Another limitation of this study is that participants

were given a wrist-worn activity monitor (LYCOS Life) to prompt them to walk every 30 minutes during one of the assessment weeks. Although the wrist monitors and EMA surveys were not programmed to coordinate with one another, it is important to note as frequent bouts of walking may be related to improved affective states [50]. Although we attempted to statistically control for the wrist-worn activity monitor by including it as a covariate in our models, this may not have entirely accounted for its impact on our findings. This is because the wrist-worn device could have influenced other unmeasured factors such as motivation, self-regulation, and social desirability [51,52]. Future studies should attempt to address this limitation.

In addition, our EMA-prompting period spanned from 7 AM to 8 PM; therefore, our findings cannot be generalized beyond these times of the day. Similarly, as our EMA-prompting schedule did not ask participants about their SBs and affective states from 8 AM to 3 PM on weekdays (because of school or summer camp schedules), the midday estimates presented were driven by weekend day data. This prompting schedule also did not allow us to stratify the models by weekend days versus weekdays. Future studies should attempt to address this limitation by prompting participants across the entire day on both weekend days and weekdays.

It is also worth noting that the EMA prompt compliance among our sample was slightly below that previously reported among other nonclinical samples of youth [53]. In addition, participants were less likely to complete EMA prompts during the morning hours, perhaps because the EMA-prompting schedule started too early in the day. To gain a better understanding of how missing data may influence the findings, models were rerun after the removal of the participant with the lowest EMA prompt compliance (approximately 33%). These results are comparable with those presented previously. Future studies with larger sample sizes should attempt to gain a better understanding of how missing data may affect study findings by stratifying analyses by participant compliance (ie, in those above vs below the average level of compliance).

The characteristics and size of our sample are also limitations of this study. For example, our participants experienced relatively low and stable levels of negative affect, limiting our ability to detect the effects of SB on negative affective states. Therefore, it is possible that some of our null findings may not be entirely because of a lack of association between SB and negative affect. However, this study warrants future confirmatory studies with larger samples, which would introduce more variability in negative affect. Power analyses for TVEMs are currently an open area of research [54]; therefore, post hoc power calculations were not completed. However, the CIs generated from TVEMs reflect the amount of data contributed at each time interval (CIs are wider on occasions when there are fewer data points) [54], aiding our understanding of the statistical power in this study. Future studies with larger samples (at the EMA prompt and person levels) are warranted as it is possible that this study was statistically underpowered, partially contributing to our null findings at some time points within the day. Studies with larger samples will also allow for the investigation of a more nuanced operationalization of the SB construct (eg, subtypes of screen-based SB) in relation to affective states across the day. A better understanding of how the different subtypes of screen-based SB relate to affective states will aid in the development of tailored intervention strategies targeting the forms of SB that appear to be most important for affective experiences. Finally, because of the time reference specified in the EMA item wording (ie, SB and affective states *right now*), this study only assessed concurrent associations.

Conclusions

This proof-of-concept study demonstrated that within-day associations between SB and affective states may not be static. Rather, these associations may differ by time of day and measurement method of SB, indicating that there may be critical windows during the day in which specific types of SB can be related to concurrent affective states. Therefore, we provide preliminary evidence to suggest that future confirmatory studies aimed at investigating the SB–affective state relationship should consider the possible time-varying nature of these associations. Future studies should also investigate the underlying variables that help explain these possible time-dependent variations in youth. We also demonstrated that self-reported screen-based SBs and device-based SB are likely distinct constructs that may be uniquely related to affective experiences. Together, we provide a preliminary justification for future investigators to carefully consider the statistical modeling and SB measurement methods they choose to use.

Acknowledgments

This work was supported by the National Institute of Diabetes and Digestive and Kidney Diseases (K01DK11306; BRB), the National Cancer Institute (T32CA009492; JZ), and the National Heart, Lung, and Blood Institute (R01HL119255; GFD).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Ecological momentary assessment prompting schedule for this study.
[DOCX File , 13 KB - formative_v6i6e37743_app1.docx]

Multimedia Appendix 2
Ecological momentary assessment survey item wording, response options, and formatting for each prompt during the assessment period.
[DOCX File , 14 KB - formative_v6i6e37743_app2.docx]

Multimedia Appendix 3
Ecological momentary assessment prompt compliance by participant characteristics (1030 prompts; N=15 participants).
[DOCX File , 14 KB - formative_v6i6e37743_app3.docx]

Multimedia Appendix 4
Frequency of ecological momentary assessment–reported screen-based sedentary behavior stratified by time of day and day of week.
[DOCX File , 13 KB - formative_v6i6e37743_app4.docx]

References

1. Tremblay MS, LeBlanc AG, Kho ME, Saunders TJ, Larouche R, Colley RC, et al. Systematic review of sedentary behaviour and health indicators in school-aged children and youth. Int J Behav Nutr Phys Act 2011 Sep 21;8:98 [FREE Full text] [doi: 10.1186/1479-5868-8-98] [Medline: 21936895]
2. Kim Y, Barreira TV, Kang M. Concurrent associations of physical activity and screen-based sedentary behavior on obesity among US adolescents: a latent class analysis. J Epidemiol 2016;26(3):137-144 [FREE Full text] [doi: 10.2188/jea.JE20150068] [Medline: 26477996]
3. Popkin BM. Urbanization, lifestyle changes and the nutrition transition. World Dev 1999 Nov;27(11):1905-1916. [doi: 10.1016/s0305-750x(99)00094-7]
4. Owen N, Sparling PB, Healy GN, Dunstan DW, Matthews CE. Sedentary behavior: emerging evidence for a new health risk. Mayo Clin Proc 2010 Dec;85(12):1138-1141 [FREE Full text] [doi: 10.4065/mcp.2010.0444] [Medline: 21123641]
5. Lenhart A, Purcell K, Smith A, Zickuhr K. Pew Internet & American Life Project. 2010 Feb 3. URL: https://files.eric.ed.gov/fulltext/ED525056.pdf [accessed 2021-11-30]
6. Salmon J, Tremblay MS, Marshall SJ, Hume C. Health risks, correlates, and interventions to reduce sedentary behavior in young people. Am J Prev Med 2011 Aug;41(2):197-206. [doi: 10.1016/j.amepre.2011.05.001] [Medline: 21767728]
7. Dunton GF, Huh J, Leventhal AM, Riggs N, Hedeker D, Spruijt-Metz D, et al. Momentary assessment of affect, physical feeling states, and physical activity in children. Health Psychol 2014 Mar;33(3):255-263 [FREE Full text] [doi: 10.1037/a0032640] [Medline: 23668846]
8. Smith KE, Haedt-Matt A, Mason TB, Wang S, Yang CH, Unick JL, et al. Associations between naturalistically assessed physical activity patterns, affect, and eating in youth with overweight and obesity. J Behav Med 2020 Dec;43(6):916-931 [FREE Full text] [doi: 10.1007/s10865-020-00152-3] [Medline: 32303944]
9. Bourke M, Hilland TA, Craike M. Contextual influences on the within-person association between physical activity and affect in adolescents: an ecological momentary assessment study. J Behav Med 2021 Jun;44(3):296-309. [doi: 10.1007/s10865-020-00197-4] [Medline: 33387176]
10. Mason TB, O'Connor SG, Schembre SM, Huh J, Chu D, Dunton GF. Momentary affect, stress coping, and food intake in mother-child dyads. Health Psychol 2019 Mar;38(3):238-247 [FREE Full text] [doi: 10.1037/hea0000714] [Medline: 30762403]
11. Hedeker D, Mermelstein RJ, Berbaum ML, Campbell RT. Modeling mood variation associated with smoking: an application of a heterogeneous mixed-effects model for analysis of ecological momentary assessment (EMA) data. Addiction 2009 Feb;104(2):297-307 [FREE Full text] [doi: 10.1111/j.1360-0443.2008.02435.x] [Medline: 19149827]
12. Larson RW, Moneta G, Richards MH, Wilson S. Continuity, stability, and change in daily emotional experience across adolescence. Child Dev 2002;73(4):1151-1165. [doi: 10.1111/1467-8624.00464] [Medline: 12146740]
13. Harding SK, Page AS, Falconer C, Cooper AR. Longitudinal changes in sedentary time and physical activity during adolescence. Int J Behav Nutr Phys Act 2015 Apr 01;12:44 [FREE Full text] [doi: 10.1186/s12966-015-0204-6] [Medline: 25888805]
14. Wen CK, Liao Y, Maher JP, Huh J, Belcher BR, Dzubur E, et al. Relationships among affective states, physical activity, and sedentary behavior in children: moderation by perceived stress. Health Psychol 2018 Oct;37(10):904-914 [FREE Full text] [doi: 10.1037/hea0000639] [Medline: 30234349]
15. Rusby JC, Westling E, Crowley R, Light JM. Psychosocial correlates of physical and sedentary activities of early adolescent youth. Health Educ Behav 2014 Feb;41(1):42-51 [FREE Full text] [doi: 10.1177/1090198113485753] [Medline: 23640122]
16. Woods HC, Scott H. #Sleepyteens: social media use in adolescence is associated with poor sleep quality, anxiety, depression and low self-esteem. J Adolesc 2016 Aug;51:41-49. [doi: 10.1016/j.adolescence.2016.05.008] [Medline: 27294324]

17. Oshima N, Nishida A, Shimodera S, Tochigi M, Ando S, Yamasaki S, et al. The suicidal feelings, self-injury, and mobile phone use after lights out in adolescents. J Pediatr Psychol 2012 Oct;37(9):1023-1030. [doi: 10.1093/jpepsy/jss072] [Medline: 22728900]
18. Dunton GF, Liao Y, Intille SS, Spruijt-Metz D, Pentz M. Investigating children's physical activity and sedentary behavior using ecological momentary assessment with mobile phones. Obesity (Silver Spring) 2011 Jun;19(6):1205-1212 [FREE Full text] [doi: 10.1038/oby.2010.302] [Medline: 21164502]
19. Yang CH, Zink J, Belcher BR, Kanning M, Dunton GF. Age-varying bi-directional associations between momentary affect and movement behaviors in children: evidence from a multi-wave ecological momentary assessment study. Ann Behav Med 2021 Aug 23;55(9):918-931 [FREE Full text] [doi: 10.1093/abm/kaaa124] [Medline: 33522571]
20. Slavish DC, Scaglione NM, Hultgren BA, Turrisi R. An ecological momentary assessment of affect, mental health symptoms, and decisions to drink among first-year college women: a pilot study. Prev Sci 2019 Jul;20(5):753-764 [FREE Full text] [doi: 10.1007/s11121-018-0966-6] [Medline: 30498934]
21. Li R, Dziak JJ, Tan X, Huang L, Wagner AT, Yang J. TVEM (Time-Varying Effect Modeling) SAS Macro Users' Guide. Version 3.1.1. The Methodology Center. 2017. URL: https://aimlab.psu.edu/wp-content/uploads/2020/02/TVEM_3.1.1-1fxcco8.pdf [accessed 2021-01-15]
22. Lanza ST, Vasilenko SA, Russell MA. Time-varying effect modeling to address new questions in behavioral research: examples in marijuana use. Psychol Addict Behav 2016 Dec;30(8):939-954 [FREE Full text] [doi: 10.1037/adb0000208] [Medline: 27736149]
23. Dunton GF, Liao Y, Dzubur E, Leventhal AM, Huh J, Gruenewald T, et al. Investigating within-day and longitudinal effects of maternal stress on children's physical activity, dietary intake, and body composition: protocol for the MATCH study. Contemp Clin Trials 2015 Jul;43:142-154 [FREE Full text] [doi: 10.1016/j.cct.2015.05.007] [Medline: 25987483]
24. Zink J, Nicolo M, Imm K, Ebrahimian S, Yu Q, Lee K, et al. Interstitial glucose and subsequent affective and physical feeling states: a pilot study combining continuous glucose monitoring and ecological momentary assessment in adolescents. J Psychosom Res 2020 Aug;135:110141 [FREE Full text] [doi: 10.1016/j.jpsychores.2020.110141] [Medline: 32447156]
25. Aminian S, Hinckson EA. Examining the validity of the ActivPAL monitor in measuring posture and ambulatory movement in children. Int J Behav Nutr Phys Act 2012 Oct 02;9:119 [FREE Full text] [doi: 10.1186/1479-5868-9-119] [Medline: 23031188]
26. Edwardson CL, Winkler EA, Bodicoat DH, Yates T, Davies MJ, Dunstan DW, et al. Considerations when using the activPAL monitor in field-based research with adult populations. J Sport Health Sci 2017 Jun;6(2):162-178 [FREE Full text] [doi: 10.1016/j.jshs.2016.02.002] [Medline: 30356601]
27. Harrington DM, Dowd KP, Bourke AK, Donnelly AE. Cross-sectional analysis of levels and patterns of objectively measured sedentary time in adolescent females. Int J Behav Nutr Phys Act 2011 Oct 28;8:120 [FREE Full text] [doi: 10.1186/1479-5868-8-120] [Medline: 22035260]
28. Saint-Maurice PF, Kim Y, Welk GJ, Gaesser GA. Kids are not little adults: what MET threshold captures sedentary behavior in children? Eur J Appl Physiol 2016 Jan;116(1):29-38. [doi: 10.1007/s00421-015-3238-1] [Medline: 26271677]
29. Do B, Mason TB, Yi L, Yang CH, Dunton GF. Momentary associations between stress and physical activity among children using ecological momentary assessment. Psychol Sport Exerc 2021 Jul;55:101935. [doi: 10.1016/j.psychsport.2021.101935] [Medline: 33841048]
30. Maras D, Flament MF, Murray M, Buchholz A, Henderson KA, Obeid N, et al. Screen time is associated with depression and anxiety in Canadian youth. Prev Med 2015 Apr;73:133-138. [doi: 10.1016/j.ypmed.2015.01.029] [Medline: 25657166]
31. Gunnell KE, Flament MF, Buchholz A, Henderson KA, Obeid N, Schubert N, et al. Examining the bidirectional relationship between physical activity, screen time, and symptoms of anxiety and depression over time during adolescence. Prev Med 2016 Jul;88:147-152. [doi: 10.1016/j.ypmed.2016.04.002] [Medline: 27090920]
32. Kuczmarski RJ, Ogden CL, Grummer-Strawn LM, Flegal KM, Guo SS, Wei R, et al. CDC growth charts: United States. Adv Data 2000 Jun 08(314):1-27. [Medline: 11183293]
33. Maher JP, Dzubur E, Huh J, Intille S, Dunton GF. Within-day time-varying associations between behavioral cognitions and physical activity in adults. J Sport Exerc Psychol 2016 Aug;38(4):423-434 [FREE Full text] [doi: 10.1123/jsep.2016-0058] [Medline: 27634288]
34. Biddle SJ, Gorely T, Marshall SJ, Cameron N. The prevalence of sedentary behavior and physical activity in leisure time: a study of Scottish adolescents using ecological momentary assessment. Prev Med 2009 Feb;48(2):151-155. [doi: 10.1016/j.ypmed.2008.10.025] [Medline: 19046984]
35. Zink J, Ebrahimian S, Belcher BR, Leventhal AM. Reciprocal associations between depression and screen-based sedentary behaviors in adolescents differ by depressive symptom dimension and screen-type. J Affect Disord 2020 Feb 15;263:39-46 [FREE Full text] [doi: 10.1016/j.jad.2019.11.130] [Medline: 31818794]
36. Warren C, Riggs N, Pentz MA. Executive function mediates prospective relationships between sleep duration and sedentary behavior in children. Prev Med 2016 Oct;91:82-88 [FREE Full text] [doi: 10.1016/j.ypmed.2016.07.024] [Medline: 27477059]

37. Zink J, Belcher BR, Imm K, Leventhal AM. The relationship between screen-based sedentary behaviors and symptoms of depression and anxiety in youth: a systematic review of moderating variables. BMC Public Health 2020 Apr 10;20(1):472 [FREE Full text] [doi: 10.1186/s12889-020-08572-1] [Medline: 32272906]
38. Lemola S, Brand S, Vogler N, Perkinson-Gloor N, Allemand M, Grob A. Habitual computer game playing at night is related to depressive symptoms. Pers Individ Differ 2011 Jul;51(2):117-122. [doi: 10.1016/j.paid.2011.03.024]
39. Lemola S, Perkinson-Gloor N, Brand S, Dewald-Kaufmann JF, Grob A. Adolescents' electronic media use at night, sleep disturbance, and depressive symptoms in the smartphone age. J Youth Adolesc 2015 Feb;44(2):405-418. [doi: 10.1007/s10964-014-0176-x] [Medline: 25204836]
40. Park SH, Kim Y. Ways of coping with excessive academic stress among Korean adolescents during leisure time. Int J Qual Stud Health Well-being 2018 Dec;13(1):1505397 [FREE Full text] [doi: 10.1080/17482631.2018.1505397] [Medline: 30099931]
41. Balantekin KN, Roemmich JN. Children's coping after psychological stress. Choices among food, physical activity, and television. Appetite 2012 Oct;59(2):298-304. [doi: 10.1016/j.appet.2012.05.016] [Medline: 22634198]
42. Coward IG. Adolescent stress: causes, consequences, and communication as an interventional model. Can J Fam Youth 2018 Mar 23;10(1):25-51. [doi: 10.29173/cjfy29341]
43. Wahab S, Rahman FN, Wan Hasan WM, Zamani IZ, Arbaiei NC, Khor SL, et al. Stressors in secondary boarding school students: association with stress, anxiety and depressive symptoms. Asia Pac Psychiatry 2013 Apr;5 Suppl 1:82-89. [doi: 10.1111/appy.12067] [Medline: 23857842]
44. Randall ET, Bohnert AM. Understanding threshold effects of organized activity involvement in adolescents: sex and family income as moderators. J Adolesc 2012 Feb;35(1):107-118. [doi: 10.1016/j.adolescence.2011.05.004] [Medline: 21752459]
45. Aparicio E, Canals J, Arija V, De Henauw S, Michels N. The role of emotion regulation in childhood obesity: implications for prevention and treatment. Nutr Res Rev 2016 Jun;29(1):17-29. [doi: 10.1017/S0954422415000153] [Medline: 27045966]
46. Snippe E, Dziak JJ, Lanza ST, Nyklíček I, Wichers M. The shape of change in perceived stress, negative affect, and stress sensitivity during mindfulness-based stress reduction. Mindfulness (N Y) 2017 Jun;8(3):728-736 [FREE Full text] [doi: 10.1007/s12671-016-0650-5] [Medline: 29057017]
47. Dziak JJ, Li R, Tan X, Shiffman S, Shiyko MP. Modeling intensive longitudinal data with mixtures of nonparametric trajectories and time-varying effects. Psychol Methods 2015 Dec;20(4):444-469 [FREE Full text] [doi: 10.1037/met0000048] [Medline: 26390169]
48. Tan X, Shiyko MP, Li R, Li Y, Dierker L. A time-varying effect model for intensive longitudinal data. Psychol Methods 2012 Mar;17(1):61-77 [FREE Full text] [doi: 10.1037/a0025814] [Medline: 22103434]
49. Zink J, Belcher BR, Dzubur E, Ke W, O'Connor S, Huh J, et al. Association between self-reported and objective activity levels by demographic factors: ecological momentary assessment study in children. JMIR Mhealth Uhealth 2018 Jun 28;6(6):e150 [FREE Full text] [doi: 10.2196/mhealth.9592] [Medline: 29954723]
50. Zink J, Berrigan DA, Broadney MM, Shareef F, Papachristopoulou A, Brady SM, et al. The effects of interrupting sitting time on affect and state anxiety in children of healthy weight and overweight: a randomized crossover trial. Pediatr Exerc Sci 2020 May 01;32(2):97-104 [FREE Full text] [doi: 10.1123/pes.2019-0169] [Medline: 32163927]
51. Casado-Robles C, Viciana J, Guijarro-Romero S, Mayorga-Vega D. Effects of consumer-wearable activity tracker-based programs on objectively measured daily physical activity and sedentary behavior among school-aged children: a systematic review and meta-analysis. Sports Med Open 2022 Jan 31;8(1):18 [FREE Full text] [doi: 10.1186/s40798-021-00407-6] [Medline: 35099630]
52. König LM, Allmeta A, Christlein N, Van Emmenis M, Sutton S. A systematic review and meta-analysis of studies of reactivity to digital in-the-moment measurement of health behaviour. Health Psychol Rev (forthcoming) 2022 Mar 21:1-25. [doi: 10.1080/17437199.2022.2047096] [Medline: 35264084]
53. Wen CK, Schneider S, Stone AA, Spruijt-Metz D. Compliance with mobile ecological momentary assessment protocols in children and adolescents: a systematic review and meta-analysis. J Med Internet Res 2017 Apr 26;19(4):e132 [FREE Full text] [doi: 10.2196/jmir.6641] [Medline: 28446418]
54. Lanza ST, Linden-Carmichael AN. Further applications and future directions. In: Lanza ST, Linden-Carmichael AN, editors. Time-Varying Effect Modeling for the Behavioral, Social, and Health Sciences. Cham, Switzerland: Springer; 2021:133-147.

Abbreviations

EMA: ecological momentary assessment
MATCH: Mothers' and Their Children's Health
SB: sedentary behavior
TVEM: time-varying effect model

Edited by A Mavragani; submitted 05.03.22; peer-reviewed by J de Brito, C Cummings; comments to author 06.04.22; revised version received 26.04.22; accepted 27.04.22; published 10.06.22.

Please cite as:
Zink J, Yang CH, Alves JM, McAlister KL, Huh J, Pentz MA, Page KA, Dunton GF, Belcher BR
Time-Varying Associations Between Device-Based and Ecological Momentary Assessment–Reported Sedentary Behaviors and the Concurrent Affective States Among Adolescents: Proof-of-Concept Study
JMIR Form Res 2022;6(6):e37743
URL: https://formative.jmir.org/2022/6/e37743
doi:10.2196/37743
PMID:35687383

Original Paper

Patient Onboarding and Engagement to Build a Digital Study After Enrollment in a Clinical Trial (TAILOR-PCI Digital Study): Intervention Study

Robert Avram[1], MSc, MD; Derek So[2], MD; Erin Iturriaga[3], MSN, DNP; Julia Byrne[4], BS; Ryan Lennon[4], MS; Vishakantha Murthy[4], MSc, MBA, PhD; Nancy Geller[3], PhD; Shaun Goodman[5,6], MSc, MD; Charanjit Rihal[4], MD; Yves Rosenberg[3], MPH, MD; Kent Bailey[4], PhD; Michael Farkouh[7], MSc, MD; Malcolm Bell[4], MD; Charles Cagin[8], DO; Ivan Chavez[9], MD; Mohammad El-Hajjar[10], MD; Wilson Ginete[11], MD; Amir Lerman[4], MD; Justin Levisay[12], MD; Kevin Marzo[13], MD; Tamim Nazif[14], MD; Jeffrey Olgin[1*], MD; Naveen Pereira[4*], MD

[1]Department of Medicine, University of California, San Francisco, San Francisco, CA, United States
[2]Department of Medicine, University of Ottawa Heart Institute, Ottawa, ON, Canada
[3]Department of Medicine, National Heart, Lung, and Blood Institute, Bethesda, MD, United States
[4]Department of Cardiovascular Medicine, Mayo Clinic Rochester, Rochester, MN, United States
[5]Department of Medicine, St. Michael's Hospital, Toronto, ON, Canada
[6]Department of Medicine, University of Alberta, Edmonton, AB, Canada
[7]Department of Medicine, University of Toronto, Toronto, ON, Canada
[8]Department of Medicine, Mayo Clinic Health System, La Crosse, WI, United States
[9]Department of Medicine, Minneapolis Heart Institute, Minneapolis, MN, United States
[10]Department of Medicine, Albany Medical College, Albany, NY, United States
[11]Department of Medicine, Essentia Institute of Rural Health, Duluth, MN, United States
[12]Department of Medicine, Northshore University Health System, Evanston, IL, United States
[13]Department of Medicine, Winthrop University Hospital, Mineola, NY, United States
[14]Department of Medicine, Columbia University Medical Center, New York, NY, United States
[*]these authors contributed equally

Corresponding Author:
Jeffrey Olgin, MD
Department of Medicine
University of California, San Francisco
505 Parnassus Avenue
San Francisco, CA, 94117
United States
Phone: 1 451 476 1325
Email: Jeffrey.Olgin@ucsf.edu

Abstract

Background: The Tailored Antiplatelet Initiation to Lessen Outcomes Due to Decreased Clopidogrel Response After Percutaneous Coronary Intervention (TAILOR-PCI) Digital Study is a novel proof-of-concept study that evaluated the feasibility of extending the TAILOR-PCI randomized controlled trial (RCT) follow-up period by using a remote digital platform.

Objective: The aim of this study is to describe patients' onboarding, engagement, and results in a digital study after enrollment in an RCT.

Methods: In this intervention study, previously enrolled TAILOR-PCI patients in the United States and Canada within 24 months of randomization were invited by letter to download the study app. Those who did not respond to the letter were contacted by phone to survey the reasons for nonparticipation. A direct-to-patient digital research platform (the Eureka Research Platform) was used to onboard patients, obtain consent, and administer activities in the digital study. The patients were asked to complete health-related surveys and digitally provide follow-up data. Our primary end points were the consent rate, the duration of

participation, and the monthly activity completion rate in the digital study. The hypothesis being tested was formulated before data collection began.

Results: After the parent trial was completed, letters were mailed to 907 eligible patients (representing 18.8% [907/4837] of total enrolled in the RCT) within 15.6 (SD 5.2) months of randomization across 24 sites. Among the 907 patients invited, 290 (32%) visited the study website and 110 (12.1%) consented—40.9% (45/110) after the letter, 33.6% (37/110) after the first phone call, and 25.5% (28/110) after the second call. Among the 47.4% (409/862) of patients who responded, 41.8% (171/409) declined to participate because of a lack of time, 31.2% (128/409) declined because of the lack of a smartphone, and 11.5% (47/409) declined because of difficulty understanding what was expected of them in the study. Patients who consented were older (aged 65.3 vs 62.5 years; P=.006) and had a lower prevalence of diabetes (19% vs 30%; P=.02) or tobacco use (6.4% vs 24.8%; P<.001). A greater proportion had bachelor's degrees (47.2% vs 25.7%; P<.001) and were more computer literate (90.5% vs 62.3% of daily internet use; P<.001) than those who did not consent. The average completion rate of the 920 available monthly electronic visits was 64.9% (SD 7.6%); there was no decrease in this rate throughout the study duration.

Conclusions: Extended follow-up after enrollment in an RCT by using a digital study was technically feasible but was limited because of the inability to contact most eligible patients or a lack of time or access to a smartphone. Among the enrolled patients, most completed the required electronic visits. Enhanced recruitment methods, such as the introduction of a digital study at the time of RCT consent, smartphone provision, and robust study support for onboarding, should be explored further.

Trial Registration: Clinicaltrails.gov NCT01742117; https://clinicaltrials.gov/ct2/show/NCT01742117

(JMIR Form Res 2022;6(6):e34080) doi:10.2196/34080

KEYWORDS

digital study; clinical trial; cardiology; smartphone; digital health; mobile health; clinical trial; mobile phone

Introduction

Background

The Tailored Antiplatelet Initiation to Lessen Outcomes Due to Decreased Clopidogrel Response After Percutaneous Coronary Intervention (TAILOR-PCI) was a large multicenter international randomized controlled trial (RCT) that compared point-of-care, genotype-guided P2Y12 inhibitor therapy to conventional clopidogrel therapy [1]. The initial follow-up duration of this trial was 1 year after the index percutaneous coronary intervention and randomization. Subsequently, the follow-up was extended to 2 years with 18- and 24-month study coordinator telephone visits. Notably, extending such follow-ups with in-person or telephone assessments of patients in large, multicenter RCTs such as the TAILOR-PCI RCT is expensive, time consuming, and complicated. The National Institutes of Health has recommended that RCTs be conducted in a pragmatic manner, including the use of digital technologies [2]. Recently, the importance of remote digital follow-up has been highlighted by the COVID-19 pandemic, during which many conventional RCTs requiring in-person recruitment and follow-up were stalled or suspended. Digital solutions to conducting RCTs provide increased convenience to both patients and enrolling sites and can potentially play a pivotal role in reducing costs and increasing accessibility to research. Given the near ubiquity of smartphones in many parts of the world and the use of mobile apps, remote RCT follow-up with regular data collection is now feasible. Whether digital technologies can be used to engage patients in a follow-up study once they are enrolled in an RCT is unknown.

The TAILOR-PCI Digital Study tested the feasibility of extending the original 1-year follow-up of the TAILOR-PCI RCT to a 2-year, remote follow-up using digital solutions and a low-contact approach (mailing letters and coordinator phone calls rather than clinic visits) to enrollment and engagement.

Objectives

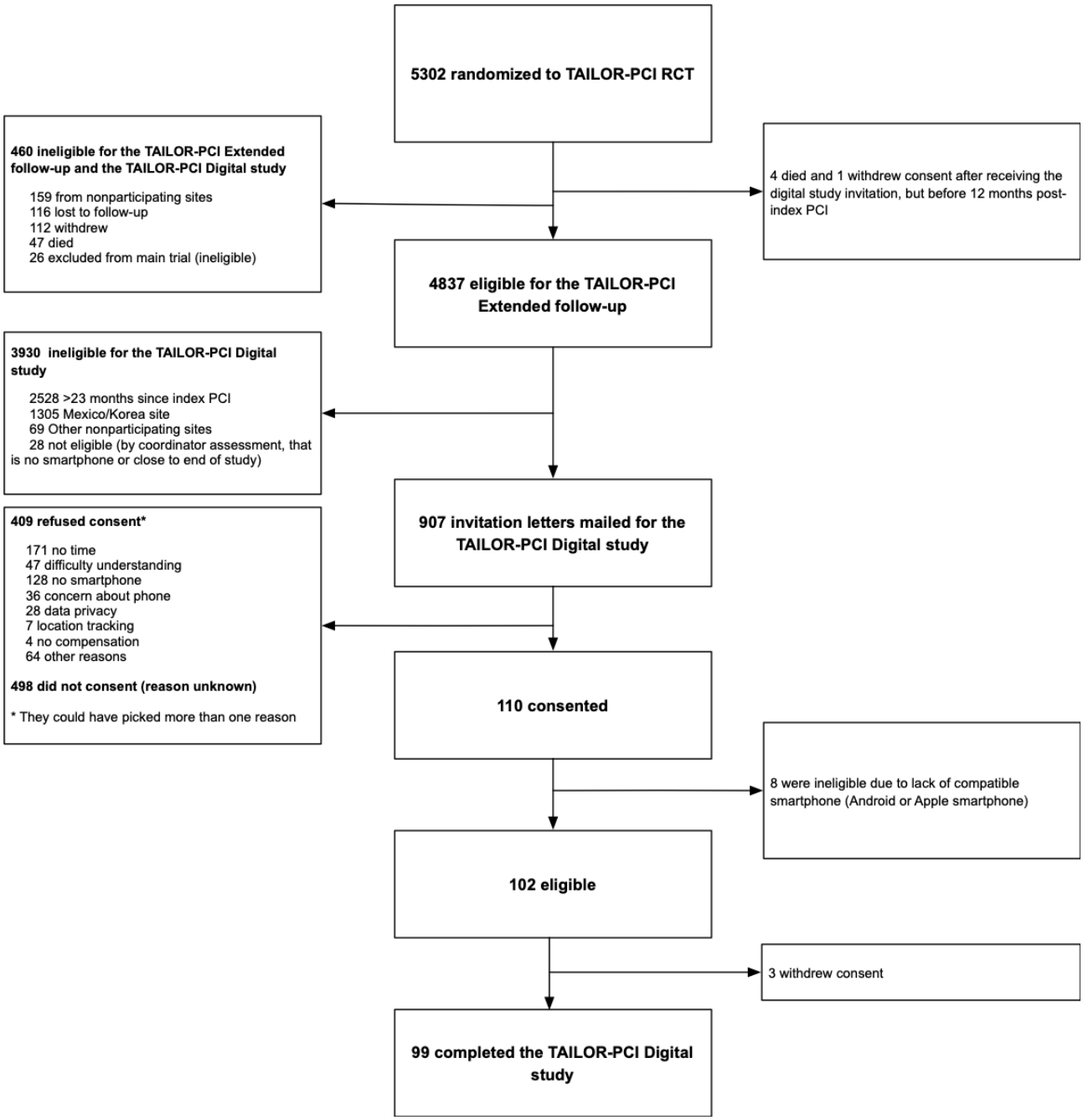
The objectives of this report are to describe our experience extending the follow-up and transitioning of the TAILOR-PCI pragmatic RCT to the TAILOR-PCI Digital Study after the main study had finished enrollment, with emphasis on patient onboarding, engagement, and results in the digital study.

Methods

Study Population

The parent TAILOR-PCI RCT (ClinicalTrials.gov NCT01742117) began enrolling patients on May 29, 2013, completed enrollment on October 31, 2018, and completed the final study follow-up a year later, with a study visit window open for up to 28 days thereafter. The TAILOR-PCI Digital Study tested the feasibility of extending RCT follow-up for up to 24 months in a subset of patients using a smartphone app designed for research. The design of the TAILOR-PCI Digital Study has been described previously [3]. Recruitment letters for the digital study were sent in February 2019. A digital study was built and conducted using the Eureka Research Platform, a direct-to-patient digital research platform [4]. TAILOR-PCI patients who enrolled from sites in the United States and Canada and were within 24 months of initial randomization and had an Apple or Android smartphone were eligible to participate (Figure 1). Of the original 34 US or Canadian TAILOR-PCI sites, 24 (71%) participated (3 declined and 7 did not have eligible patients).

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) diagram. PCI: percutaneous coronary intervention; RCT: randomized controlled trial; TAILOR-PCI: Tailored Antiplatelet Initiation to Lessen Outcomes Due to Decreased Clopidogrel Response After Percutaneous Coronary Intervention.



Recruitment

Recruitment was initiated by TAILOR-PCI site study coordinators who mailed letters to eligible patients and invited them to participate. Patients were instructed to visit the study website to learn more about the digital study, read and sign the consent form (if they chose to participate), and then receive an SMS text message with a link to download the study mobile app (Figure S1 in Multimedia Appendix 1). Embedded in the invitation letter was a unique patient code, a patient-specific *one-time access code* to establish 1:1 linkage with the TAILOR-PCI study ID, allowing synchronization of the data collected through Eureka and the RCT. Those who did not consent to the digital study after receiving the initial invitation letter were contacted via telephone by site coordinators and were asked whether they wanted additional information on the digital study or help with the smartphone app installation process and were encouraged to participate. Reasons for not participating were elicited (Multimedia Appendix 1), and data on the patients' education level and computer literacy were also obtained.

Oversight

The Mayo Clinic was the clinical coordinating center for all participating sites, and the University of California, San Francisco (UCSF) was the digital technology center. The UCSF developed the digital component of the study and provided technical support to the coordinating center throughout the study period. The Mayo Clinic conceived the study, received

institutional review board (IRB) approval, and operationalized the implementation of the digital study, whereas the UCSF received IRB approval for the Eureka Research Platform as a digital coordinating center. Each participating eligible TAILOR-PCI site obtained local IRB approval for the study invitation material and for making patient contact. An independent National Heart, Lung, and Blood Institute–appointed Observational Study Monitoring Board was responsible for overseeing the conduct, safety, and data of the study.

Ethics Approval

This study was approved by the Mayo Clinic IRB (number: 11-006837). The Eureka Platform used to conduct this study was approved by the UCSF IRB (number: 17-21879).

Data Collection

At baseline, patients completed the following patient-reported outcomes (PROs) instruments on the mobile app: Duke Activity Score Index [5], Seattle Angina Questionnaire (SAQ) [6], and Modified Medical Research Council Dyspnea Scale [7] (Figure S1 in Multimedia Appendix 1). Patients also entered their medications with dosages using the Eureka medication tool [8]. These activities were repeated every month. The patients also completed a weekly 2-question angina diary. Anxiety scores were collected at baseline and every 6 months using the General Anxiety Disorder 7-item (GAD-7) questionnaire [9]. If monthly activities were not completed, patients received weekly automated SMS text messages and push notifications to remind them to complete the study activities.

Aim of This Study

The main aim of the digital study was to determine the feasibility of transitioning a clinical trial to a digital study, and the aims were defined previously [3]. First, we described the proportion of patients who were invited to participate in the digital follow-up and consented. We further compared the patient characteristics of those who consented to participate with those who were eligible but did not consent. Second, among the consented patients, we measured the duration of participation in the study (duration in months between the first and last digital study activity completed by the patient). Third, we measured the proportion of enrolled patients who participated in at least 80% of the eligible digital visits (the total number of visits varied between 1 and 24 according to the date of enrollment in the main RCT). Additional outcomes that were measured were the proportion of consented patients who downloaded the Eureka app, average time until the study drop-off (described as skipping ≥1 month of activities and not re-engaging with the Eureka app despite Mayo Clinic digital study coordinator phone calls), digital visit completion rate (number of monthly digital visits completed over the number of visits available), attrition rate (1 minus the digital visit completion rate), and activity completion rate (number of activities completed over the number of activities available, stratified by weekly, monthly, and biyearly activities). Finally, as an exploratory aim, PROs collected in the digital study were described and stratified by the randomization arm in the parent trial.

Data Analysis

The Duke Activity Score Index, Modified Medical Research Council Dyspnea Scale, SAQ, and GAD-7 were scored according to their respective instructions [5,7,9]. Continuous variables are presented as “mean (SD)” if approximately symmetrically distributed and as “median (IQR)” otherwise and were compared using the *t* test (2-tailed) or the Mann–Whitney *U* test as appropriate. Categorical variables were presented as “frequency (percentage)” and compared using either chi-square or Fisher exact tests, and 2-tailed *P* values <.05 were considered statistically significant, without further correction for multiple testing. Binary outcomes were reported with 95% CIs for the percentage using the Agresti-Coull method for interval estimation [10]. CIs for continuous variables were estimated using normal approximations for the mean, using transformations as needed. The digital visit participation rate and survey completion rates were calculated within participants to estimate a percentage and then summarized across individuals as continuous measures. Univariate logistic regression models were used to determine the effects of baseline characteristics (age, sex, etc) on the likelihood of consenting to the digital study. Significant variables in the univariate analysis ($P \leq .05$) were included in this exploratory multivariable logistic regression model to identify the association between the different variables and the likelihood of consenting to the digital study. We present the activity results stratified by the randomization arm in the TAILOR-PCI RCT and grouped by bins of time since the index procedure. We assessed the survey results over time using a mixed model approach, allowing for random effects for intercept and slope within participants and modeled an overall intercept and slope as fixed effects, as well as with interaction between the randomized arm (genotype-guided vs conventional) and the slope. Data were analyzed using Python 3.5 and SAS (version 9.4; SAS Institute).

Results

Participation

Letters were mailed to 907 patients across 24 eligible sites in the United States and Canada (Figure 1), who had completed an average follow-up of 15.6 (SD 5.2; median 16.8, IQR 11.0-20.2) months since randomization in the parent RCT. In all, 2 sites mailed letters and had no patients enrolled in the digital study. These letters led to 31.9% (290/907 invited patients) study information webpage visits and 13.3% (121/907) registrations. A total of 12.1% (110/907) patients consented, among whom 92.7% (102/907) were eligible.

Consent for the TAILOR-PCI Digital Study and Patient Characteristics

Of the 110 patients who consented, 45 (40.9%) did so after the invitation letter alone, whereas 37 (33.6%) and 28 (25.5%) consented after the first and second calls, respectively. The median time from randomization to invitation by letter of those who consented to the digital study as compared with those that did not consent was not different (median 17, IQR 5-23 vs median 17, IQR 4-24 months; $P=.47$). The mean age of the consented patients was 65.3 (SD 9.0) years versus 62.5 (SD 11.0) years for nonconsented patients ($P=.006$), and most of

those who consented were male (91/110, 82.7% vs 594/797, 74.8%; P=.06). Comorbidities were similar among those who consented and those who did not, except cigarette use (7/110, 6.4% vs 198/797, 24.8%; P<.001) and diabetes (21/110, 19.1% vs 237/797, 29.7%; P=.02), which were less prevalent in the consented group (Table 1). A greater proportion of consenting patients had a bachelor's degree or higher (75/106, 70.8% vs 233/620, 37.6%; P<.001). Among those who consented to participate in the education and computer literacy questionnaire, we observed a higher proportion of daily internet use (96/106, 90.5% vs 389/620, 62.7%; P<.001) and smartphone ownership (99/101, 98% vs 397/462, 86%; P<.001; Table 1).

Table 1. Baseline characteristics.

Variable	Consented patients (N=110)	Eligible patients (no consent) (N=797)	*P* value[a]
Hospital presentation randomization group, n (%)			.62
Stable coronary artery disease	33 (30)	214 (26.8)	
Unstable angina or non-STEMI[b]	57 (51.8)	409 (51.3)	
STEMI	20 (18.2)	174 (21.8)	
Sex, n (%)			.06
Male	91 (82.7)	594 (74.5)	
Female	19 (17.2)	203 (25.5)	
Age (years)			
At randomization			.006
Value, mean (SD)	65.3 (9.0)	62.5 (11.0)	
Value, median (IQR)	65 (47-87)	62 (28-95)	
Men			<.001
Value, mean (SD)	65.6 (8.9)	61.2 (10.5)	
Value, median (IQR)	65 (47-87)	61 (28-95)	
Women			.38
Value, mean (SD)	64.1 (9.5)	66.3 (11.7)	
Value, median (IQR)	65 (47-81)	68 (36-95)	
Ethnicity, n (%)			.19
White	99 (90)	657 (82.4)	
Asian	3 (2.7)	27 (3.4)	
African American	1 (0.9)	37 (4.6)	
Hispanic or Latino	0 (0)	14 (1.7)	
Other	7 (6.4)	62 (7.8)	
Country, n (%)			.61
Canada	23 (20.9)	184 (23.1)	
United States	87 (79.1)	613 (76.9)	
BMI (kg/m^2), n (%)			.18
<25	19 (17.3)	124 (15.6)	
25-30	49 (44.5)	307 (38.5)	
≥30	42 (38.2)	363 (45.5)	
Diabetes, n (%)	21 (19.1)	237 (29.7)	.02
Hypertension, n (%)	74 (67.2)	554 (69.5)	.63
Dyslipidemia, n (%)	77 (70)	540 (67.8)	.64
Any history of heart failure, n (%)	2 (1.8)	28 (3.5)	.35
Heart failure >2 weeks, n (%)	1 (0.9)	24 (3)	.21
Estimated glomerular filtration rate (modification of diet in renal disease) <60, n (%)	14 (12.7)	91 (11.4)	.61
Cigarette use, n (%)	7 (6.4)	198 (24.8)	<.001
History of myocardial infarction (excluding index event), n (%)	14 (12.7)	136 (17.1)	.25
Peripheral artery disease, n (%)	3 (2.7)	33 (4.1)	.48
History of percutaneous coronary intervention, n (%)	27 (24.5)	220 (27.6)	.50

Variable	Consented patients (N=110)	Eligible patients (no consent) (N=797)	*P* value[a]
History of coronary artery bypass grafting, n (%)	11 (10)	75 (9.4)	.84
Stroke or transient ischemic attack, n (%)	2 (1.8)	26 (3.3)	.41
Family history of coronary artery disease, n (%)	61 (55.4)	413 (51.8)	.48
Chronic lung disease, n (%)	4 (3.6)	42 (5.3)	.46
Currently on dialysis, n (%)	0 (0)	1 (0.1)	.71
Education and computer literacy form completed, n (%)			<.001
Completed the form	106 (96.4)	637 (79.9)	
Did not complete the form	4 (3.6)	160 (20.1)	
Education level			<.001
Less than high school	2 (1.9)	39 (6.1)	
High school graduate or some college	25 (23.5)	274 (43)	
Associate or bachelor's degree	50 (47.2)	164 (25.7)	
Graduate or PhD	25 (23.6)	75 (11.7)	
Prefer not to answer	4 (3.8)	85 (13.4)	
Frequency of internet use			<.001
Does not use	2 (1.9)	76 (11.8)	
About daily	96 (90.5)	397 (62.3)	
About once a week	2 (1.9)	46 (7.2)	
Occasionally (less than once a week)	5 (4.7)	44 (6.9)	
Do not know	0 (0)	2 (0)	
Prefer not to answer	1 (0.9)	72 (11.3)	
Has a computer or laptop	95 (93)	423 (88)	.17
Has a smartphone	99 (98)	404 (86)	<.001
Has a tablet	45 (49)	204 (46)	.63
Has a smart speaker	26 (31)	74 (19)	.02
Has downloaded app to phone	93 (94)	326 (81)	.006

[a]Comparison of consented and nonconsented patients.

[b]STEMI: ST-elevation myocardial infarction.

Nonparticipation in the Digital Study: Reasons and Predictors

Among the 409 patients surveyed during phone follow-up regarding reasons for nonparticipation in the digital study, the most common reasons reported (Table 2) were lack of time (171/409, 41.8%), lack of smartphone use (128/409, 31.3%), and difficulty in understanding what was expected of them in the study (47/409, 11.5%). Concerns about data privacy (28/409, 6.9%) and location tracking (7/409, 1.7%) were less frequent.

Multivariable analysis revealed that older age, higher educational level, daily internet use, nonsmoking status, and nondiabetic status were significant independent predictors of consenting to the digital study (Table S1 in Multimedia Appendix 1). A sex–age interaction was observed, such that women aged 50 years were more likely than men at the same age to consent (odds ratio 1.55, 95% CI 0.58-4.16), but women aged ≥70 years were less likely to consent (odds ratio 0.46, 95% CI 0.24-0.87; interaction *P*=.03; Table S1 in Multimedia Appendix 1).

Table 2. Reasons given for declining participation in the digital follow-up.

Variable[a]	Overall participants (N=409), n (%)
No time	171 (41.8)
No smartphone	128 (31.3)
Difficulty in understanding	47 (11.5)
Concern about phone	36 (8.8)
Data privacy	28 (6.8)
No compensation	4 (1)
Location tracking	7 (1.7)
Other reasons	64 (15.6)

[a]Patients could choose more than 1 reason for not participating in digital follow-up.

Engagement and Completion of Digital Visits and Activities

The median duration of follow-up for the digital study was 6.9 (IQR 3.0-12.3) months and patients participated for a median of 5.3 (IQR 2.2-10.9) months before dropping off. Until the date of the last follow-up, of the 102 consented and eligible participants, 65 (63.7%) remained engaged in the digital study, 61 (59.8%) patients completed ≥80%, and 41 (40.2%) completed <80% of all available digital visits (collection of surveys at one point in time). Among the 920 monthly digital visits made available to the patients, 577 (62.7%) were fully completed, 120 (13%) were partially completed, and 223 (24.2%) were skipped. The participation rate for the study e-Visits was constant throughout the course of the digital study (Figure 2), with patients completing 64.9% (SD 7.6%) of activities presented to them. A total of 55.48% (3525/6354) available activities were completed by patients, and this proportion increased to 67.06% (2443/3643) when the weekly angina diary was excluded (Figure 3). Out of the eligible patients, the completion rate of activities was 76.2% (78/102) between baseline and month 4, 62.1% (29/47) between month 5 and month 9, 59% (13/22) between month 10 and month 14, and 62.3% (9/14) between month 15 and month 20.

Figure 2. Tailored Antiplatelet Initiation to Lessen Outcomes Due to Decreased Clopidogrel Response After Percutaneous Coronary Intervention e-visit completion rate.

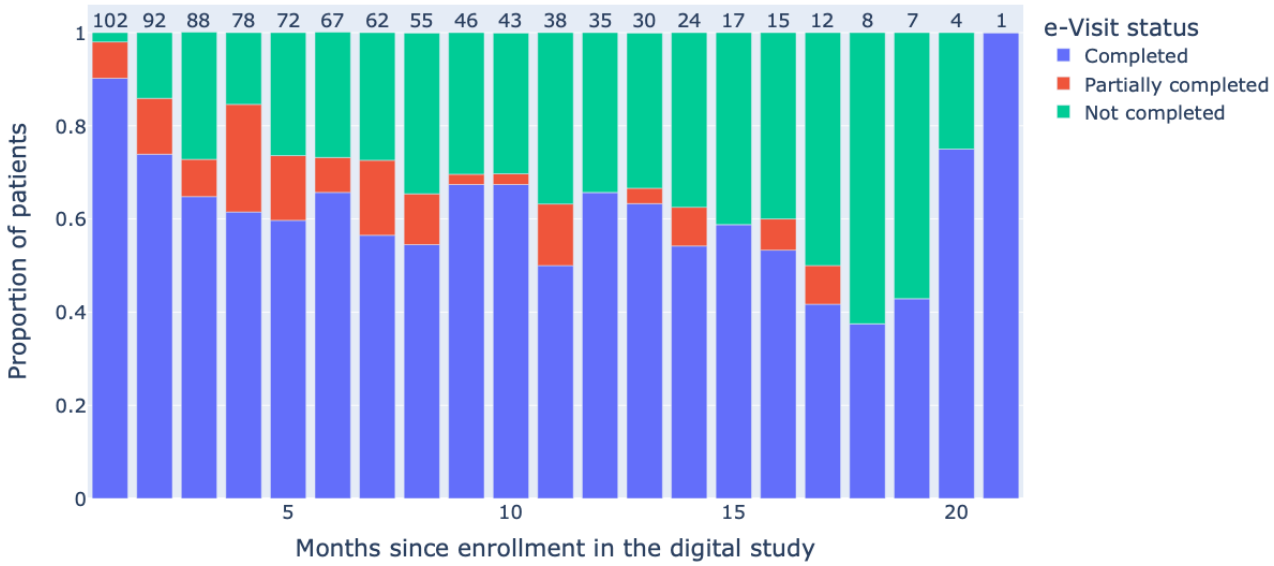
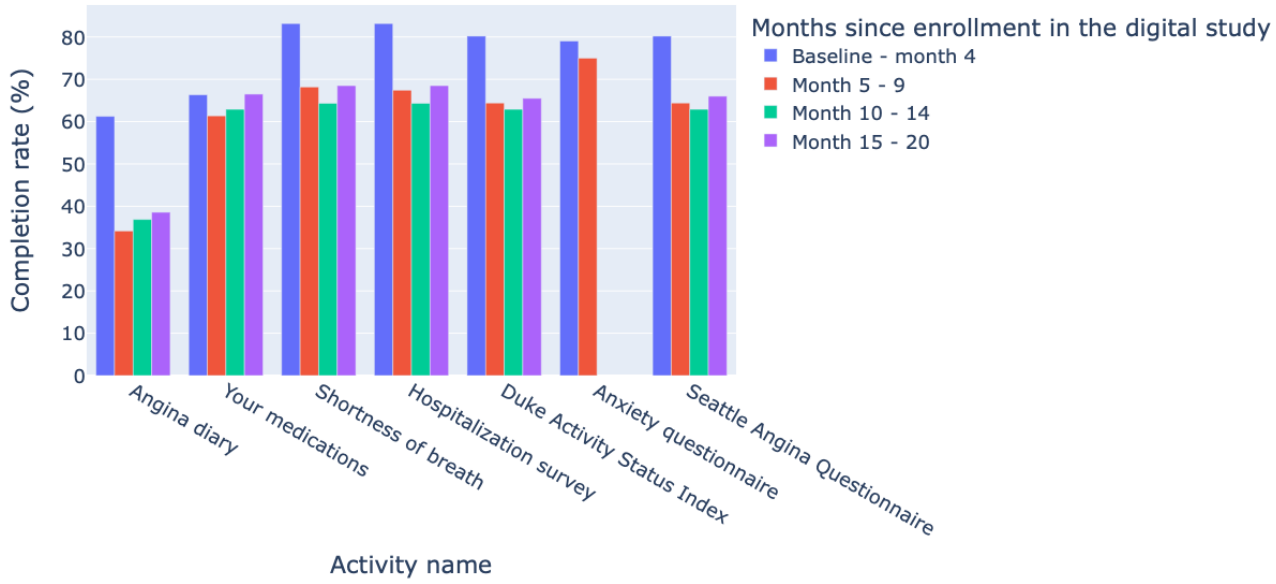


Figure 3. Tailored Antiplatelet Initiation to Lessen Outcomes Due to Decreased Clopidogrel Response After Percutaneous Coronary Intervention activities completion rates.



PRO Findings

There were no differences among the randomization arms in any PROs (Figure S2 in Multimedia Appendix 1). There was a significant drop of >10 points in the SAQ disease perception (about 1.0 points less per 4 weeks; P=.02), but not in the other SAQ subdomains or in the overall SAQ score between the start of the study and the end of the digital study (Figure S2 and Table S2 in Multimedia Appendix 1). The patient Duke Activity Status Index trended lower throughout the study (approximately 0.4 points less per 4 weeks; P=.004; Figure S2 and Table S2 in Multimedia Appendix 1).

Discussion

Principal Findings

The TAILOR-PCI Digital Study was a proof-of-concept study that attempted to extend the follow-up of the main TAILOR-PCI RCT by using a digital platform at the end of the parent trial once enrollment was completed. The modest rate of participation at 12.1% (110/907) was limited by patients not visiting the home page after being mailed an invitation letter. Among those who responded to the invitation (171/409, 41.8%), perceived lack of time and lack of a smartphone were the most common reasons for declining participation.

There may be several reasons for the low participation rate in this proof-of-concept study. First, enrollment in the digital study began after completion of enrollment in the parent trial. The digital study was an *add-on* and was not integrated during initial recruitment in the parent RCT. Second, recruitment in the digital study used a low-touch approach with mailing letters and a limited number of phone call attempts (3 in total). Third, as email addresses were not collected in the parent trial, no email invitations were sent, and we relied on mailed paper letters for invitation to the digital study. There was a significant patient drop-off between the initial mail-in invitation to join the digital study and visits to the study webpage. One can speculate that the visit rate to the study webpage could have been increased by using email or SMS text message invitations and a higher-touch recruitment method, such as in-person enrollment during initial enrollment or subsequent follow-up in the parent trial.

Among those who consented (110/907, 12.1%), the engagement rate was excellent for a digital study that had minimal coordinator interaction, with 60% (66/110) of the patients completing 80% of the digital monthly visits. The digital study also demonstrated that once patients consented, the collection of a large amount (3525 digital PRO forms) and a wide variety of data, including a weekly angina diary, medications, the SAQ, and GAD-7, was feasible. This type of data can be used to phenotype patients enrolled in an RCT using digital technology.

Reasons for High Digital Engagement and Retention Rates

The engagement and retention rates observed in our study among consented patients were higher than those reported in 12 previously described digital studies [11]. We did not observe an increase in the attrition rate during the duration of the study. This is in contrast to the median duration of participation of 5.5 days across several large-scale digital studies [11-15]. In these studies, only a fraction of patients contributed data from days 29 to 50 of recruitment [11-15]. For example, in a large asthma study, among 6470 consented patients, only 175 completed the required 6-month follow-up [13]. Several factors may have played a role in the higher engagement and retention rates observed in the TAILOR-PCI Digital Study. First, patients were already part of the TAILOR-PCI RCT and had maintained a 1-year clinical follow-up and therefore could represent patients who are more familiar with research protocols and more compliant with follow-up. Second, they experienced coronary artery disease, which is the clinical condition of interest studied in the digital study. Recruiting patients to a study that researches their medical conditions has previously been described as a factor of increased retention in digital studies [11,16]. Third,

transparency in disclosing details of study participation, such as study duration, monthly time commitment, and data being collected, may have preselected motivated patients but could also have contributed to the lower than anticipated consent rate. Fourth, the observed lower attrition rate could also be owing to the several strategies used in the digital study to maximize engagement. Eureka has a robust automated messaging and reminder system that notifies patients when new activities are available or if they have not been completed.

Challenges to Overcome and Possible Solutions to Implement Digital Technology in RCTs

Although the consent rate was lower than anticipated in our study, the 32% website visit rate observed in this study was higher than the 1% to 3% website visit rate reported in marketing campaigns that use emails to direct users to websites [17] or the 0.8% response rate for a large nationwide trial that used email invitations [18]. Successful recruitment for digital studies typically requires massive social media campaigns and a large number of invitations directed to a reasonable number of patients [11]. Moreover, experience with other studies on the Eureka platform and other digital studies suggests that higher-touch enrollment involving multiple phone call attempts or in-person enrollment during a study visit where technical concerns can be addressed at the time of digital enrollment are more effective [19], particularly in older populations such as those in TAILOR-PCI.

We identified several major challenges that should be addressed in future digital studies to increase consent rates. The digital study was designed and launched after enrollment in the main RCT was completed, thus requiring separate consent to be obtained months after the initial RCT enrollment. Therefore, we speculate that including a digital component at the inception of an RCT or at the time of initial consent may not only increase consent for digital studies but may also improve consent rates for RCTs by enabling easier follow-up. The lack of in-person visits with a study coordinator at the time of enrollment may have deterred the engagement of patients who had or perceived technical difficulties. For instance, among those eligible, 20% (1/5) of patients who did not participate in the digital study had never previously downloaded a smartphone app. A previous study demonstrated that in-clinic recruitment, as opposed to low-touch self-enrollment in a digital study, can increase both consent rate and engagement [11].

The lack of a smartphone was a major barrier to participating in the digital study, yet the prevalence of smartphone use, as reported by site study coordinators, was high at 75.8% (97/128) among those who cited it as a reason not to participate. This is comparable with the prevalence of smartphone use among those aged >65 years [20]. Although not directly explored in our study, patients may have lacked a compatible Android or Apple smartphone, may not have had a data plan, or may have been worried about the cost of data use. Strategies such as the provision of study-specific mobile devices or data plans to patients could enable easier implementation of digital technology in RCTs. Only 6.8% (28/409) of the patients who did not consent cited data privacy as a potential obstacle. We observed that data privacy concerns are often brought forward by IRBs and physicians as a potential barrier to digital studies, whereas patients themselves are less concerned [21]. Furthermore, the Eureka digital platform, which has engaged >400,000 study patients across 45 studies, is affiliated with an academic institution, the University of California San Francisco, which may inspire increased trust compared with commercial entities [22]. Finally, involving patients in the design and conduct of a digital study, which was not done for this digital study, could provide substantial value and lead to higher recruitment and engagement [23].

Limitations

Our study has several limitations. Only a fraction of patients (110/907, 12.1%) from the main RCT consented for the digital study; therefore, our conclusions cannot be generalized to the parent trial cohort. Moreover, patients who consented to the digital study were not representative of those enrolled in the main RCT, as they were predominantly highly educated, healthier (did not smoke or have diabetes), of White ethnicity, and technologically literate. Although this disparity may affect RCTs, especially digital studies [24,25], specific efforts must be made to increase the recruitment of underrepresented minorities in digital studies. Strategies such as the inclusion of a smartphone with a data plan for eligible patients could reduce the accessibility barrier and possibly enroll more diverse populations. Although our study experienced higher engagement rates than other digital studies, dropout rates observed in digital studies were higher than those in standard RCTs (but not necessarily higher than follow-up registries after completion of the RCT follow-up); therefore, appropriate statistical considerations need to be given, including novel trial designs specifically applicable to mobile health studies. Finally, because of the lower than anticipated consent rate, we were able to obtain clinical information digitally only from a small subset of patients enrolled in the main RCT; therefore, our clinical findings are descriptive and hypothesis-generating.

Conclusions

Extended follow-up of the TAILOR-PCI RCT using a digital platform was technically feasible; however, enrollment and consent rates in this study population were significantly limited. Once enrolled in the digital study, engagement was initially high, but the digital activity completion rate was modest. The reasons for low enrollment and modest activity completion rate by patients using this digital technology deserve further exploration.

Acknowledgments

This work was funded by a grant from the National Heart, Lung, and Blood Institute–National Institutes of Health (principal investigators: NP and JO [grant 3U01HL128606-03S1]; principal investigator: NP [grant U01 HL128606]; principal investigator

JO [grant U2CEB021881]). RA is supported by the *Fonds de la recherche en santé du Québec* (grant 274831). The authors are grateful for the effort and assistance from Daniel Crusan, BS; Monica Olson, MBA, Mayo Clinic; and the Tailored Antiplatelet Initiation to Lessen Outcomes Due to Decreased Clopidogrel Response After Percutaneous Coronary Intervention clinical trial site principal investigators and their study coordinators. The Tailored Antiplatelet Initiation to Lessen Outcomes Due to Decreased Clopidogrel Response After Percutaneous Coronary Intervention clinical trial site principal investigators include J Dawn Abbott, MD, Rhode Island Hospital; Khaled Abdul-Nour, MD, Henry Ford Health System; William Campbell, MD, University of Mississippi Medical Center; Amir Darki, MD, Loyola University Medical Center; Josh Doll, MD, Greenville Health System; Vlad Dzavik, MD, Toronto General Hospital; Paul Gordon, MD, The Miriam Hospital; John Graham, MD, St Michael's Hospital; Louie Kostopoulos, MD, Aurora Research Institute; Gary Lane, MD, Mayo Clinic in Florida; Mina Madan, MD, Sunnybrook Research Institute; D Fearghas O'Cochlain, MD, Mayo Clinic Health System—Eau Claire Hospital; Louai Razzouk, MD, New York University Langone Medical Center; John Sweeney, MD, Mayo Clinic in Arizona; and Steven Weitz, MD, Cardiology Associates of Schenectady. None of these individuals received compensation for their participation in the study. The content of this manuscript is solely the responsibility of the authors and does not necessarily reflect the views of the National Heart, Lung, and Blood Institute–National Institutes of Health or United States Department of Health and Human Services.

Authors' Contributions

RA, JO, NP, JB, and RL analyzed and interpreted the data. JB conducted follow-up phone calls with study coordinators across different recruitment sites. RA, NP, and JO drafted the manuscript. VM helped with the figure generation. All authors revised the content of the manuscript, contributed significantly to its improvement, and read and approved its final version.

Conflicts of Interest

SG has received research grant support and honoraria from: Amgen, Anthos Therapeutics, AstraZeneca, Bayer, Boehringer Ingelheim, Bristol Myers Squibb, CSL Behring, Daiichi-Sankyo/American Regent, Eli Lilly, Esperion, Ferring Pharmaceuticals, HLS Therapeutics, JAMP Pharma, Merck, Novartis, Novo Nordisk A/C, Pendopharm/Pharmascience, Pfizer, Regeneron, Sanofi, Servier, Valeo Pharma; and salary support/honoraria from the Heart and Stroke Foundation of Ontario/University of Toronto (Polo) Chair, Canadian Heart Research Centre and MD Primer, Canadian VIGOUR Centre, Cleveland Clinic Coordinating Centre for Clinical Research, Duke Clinical Research Institute, New York University Clinical Coordinating Centre, PERFUSE Research Institute, and TIMI Study Group (Brigham Health). MF has received research grant support from Amgen, Astra Zeneca, Novartis, and Novo Nordisk. .

Multimedia Appendix 1
Digital study.
[DOCX File , 991 KB - formative_v6i6e34080_app1.docx]

References

1. Pereira NL, Farkouh ME, So D, Lennon R, Geller N, Mathew V, et al. Effect of genotype-guided oral P2Y12 inhibitor selection vs conventional clopidogrel therapy on ischemic outcomes after percutaneous coronary intervention: the TAILOR-PCI randomized clinical trial. JAMA 2020 Aug 25;324(8):761-771 [FREE Full text] [doi: 10.1001/jama.2020.12443] [Medline: 32840598]
2. The NIH collaboratory launches a new resource on methods and best practices for pragmatic clinical trials. National Center for Complementary and Integrative Health, National Institutes of Health. 2019 Oct 03. URL: https://nccih.nih.gov/research/blog/collaboratory-living-textbook [accessed 2020-11-16]
3. Pereira NL, Avram R, So DY, Iturriaga E, Byrne J, Lennon RJ, et al. Rationale and design of the TAILOR-PCI digital study: transitioning a randomized controlled trial to a digital registry. Am Heart J 2021 Feb;232:84-93 [FREE Full text] [doi: 10.1016/j.ahj.2020.10.069] [Medline: 33129990]
4. UCSF Eureka Team. Eureka – a digital research platform. University of California San Francisco. URL: https://info.eurekaplatform.org/ [accessed 2022-03-23]
5. Hlatky MA, Boineau RE, Higginbotham MB, Lee KL, Mark DB, Califf RM, et al. A brief self-administered questionnaire to determine functional capacity (the Duke Activity Status Index). Am J Cardiol 1989 Sep 15;64(10):651-654. [doi: 10.1016/0002-9149(89)90496-7] [Medline: 2782256]
6. Spertus JA, Winder JA, Dewhurst TA, Deyo RA, Prodzinski J, McDonell M, et al. Development and evaluation of the Seattle Angina Questionnaire: a new functional status measure for coronary artery disease. J Am Coll Cardiol 1995 Feb;25(2):333-341 [FREE Full text] [doi: 10.1016/0735-1097(94)00397-9] [Medline: 7829785]
7. Mahler DA, Wells CK. Evaluation of clinical methods for rating dyspnea. Chest 1988 Mar;93(3):580-586. [doi: 10.1378/chest.93.3.580] [Medline: 3342669]
8. Bell DS, O'Neill SM, Reynolds K, Schoeff D. Evaluation of RxNorm in ambulatory electronic prescribing. Rand Health Q 2011;1(3):8 [FREE Full text] [Medline: 28083195]

9. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
10. Agresti A, Coull BA. Approximate is better than "exact" for interval estimation of binomial proportions. Am Stat 1998 May;52(2):119-126. [doi: 10.2307/2685469]
11. Pratap A, Neto EC, Snyder P, Stepnowsky C, Elhadad N, Grant D, et al. Indicators of retention in remote digital health studies: a cross-study evaluation of 100,000 participants. NPJ Digit Med 2020;3:21 [FREE Full text] [doi: 10.1038/s41746-020-0224-8] [Medline: 32128451]
12. Bot BM, Suver C, Neto EC, Kellen M, Klein A, Bare C, et al. The mPower study, Parkinson disease mobile data collected using ResearchKit. Sci Data 2016 Mar 03;3:160011 [FREE Full text] [doi: 10.1038/sdata.2016.11] [Medline: 26938265]
13. Chan YY, Wang P, Rogers L, Tignor N, Zweig M, Hershman SG, et al. The Asthma Mobile Health Study, a large-scale clinical observational study using ResearchKit. Nat Biotechnol 2017 Apr;35(4):354-362 [FREE Full text] [doi: 10.1038/nbt.3826] [Medline: 28288104]
14. Hershman SG, Bot BM, Shcherbina A, Doerr M, Moayedi Y, Pavlovic A, et al. Physical activity, sleep and cardiovascular health data for 50,000 individuals from the MyHeart Counts Study. Sci Data 2019 Apr 11;6(1):24 [FREE Full text] [doi: 10.1038/s41597-019-0016-7] [Medline: 30975992]
15. Perez MV, Mahaffey KW, Hedlin H, Rumsfeld JS, Garcia A, Ferris T, Apple Heart Study Investigators. Large-scale assessment of a smartwatch to identify atrial fibrillation. N Engl J Med 2019 Nov 14;381(20):1909-1917 [FREE Full text] [doi: 10.1056/NEJMoa1901183] [Medline: 31722151]
16. Chaibub Neto E, Pratap A, Perumal TM, Tummalacherla M, Snyder P, Bot BM, et al. Detecting the impact of subject characteristics on machine learning-based diagnostic applications. NPJ Digit Med 2019;2:99 [FREE Full text] [doi: 10.1038/s41746-019-0178-x] [Medline: 31633058]
17. Berry S. What is a good conversion rate? (and how does yours compare?). WebFX. 2020 Apr 15. URL: https://www.webfx.com/blog/marketing/what-is-a-good-conversion-rate/ [accessed 2022-03-23]
18. Baca-Motes K, Edwards AM, Waalen J, Edmonds S, Mehta RR, Ariniello L, et al. Digital recruitment and enrollment in a remote nationwide trial of screening for undiagnosed atrial fibrillation: lessons from the randomized, controlled mSToPS trial. Contemp Clin Trials Commun 2019 Jun;14:100318 [FREE Full text] [doi: 10.1016/j.conctc.2019.100318] [Medline: 30656241]
19. Spartano NL, Lin H, Sun F, Lunetta KL, Trinquart L, Valentino M, et al. Comparison of on-site versus remote mobile device support in the Framingham heart study using the health eHeart study for digital follow-up: randomized pilot study set within an observational study design. JMIR Mhealth Uhealth 2019 Sep 30;7(9):e13238 [FREE Full text] [doi: 10.2196/13238] [Medline: 31573928]
20. Smartphone ownership in advanced economies higher than in emerging. Pew Research Center. 2019 Feb 04. URL: https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/pg_global-technology-use-2018_2019-02-05_0-01/ [accessed 2022-03-23]
21. Avram R, Marquis-Gravel G, Simard F, Pacheco C, Couture É, Tremblay-Gravel M, et al. Understanding the patient perspective on research access to national health records databases for conduct of randomized registry trials. Int J Cardiol 2018 Jul 01;262:110-116. [doi: 10.1016/j.ijcard.2017.12.074] [Medline: 29706388]
22. Pratap A, Allred R, Duffy J, Rivera D, Lee HS, Renn BN, et al. Contemporary views of research participant willingness to participate and share digital data in biomedical research. JAMA Netw Open 2019 Nov 01;2(11):e1915717 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.15717] [Medline: 31747031]
23. Faulkner M, Alikhaani J, Brown L, Cruz H, Davidson D, Gregoire K, et al. Exploring meaningful patient engagement in ADAPTABLE (aspirin dosing: a patient-centric trial assessing benefits and long-term effectiveness). Med Care 2018 Oct;56 Suppl 10 Suppl 1:S11-S15 [FREE Full text] [doi: 10.1097/MLR.0000000000000949] [Medline: 30074943]
24. Inan OT, Tenaerts P, Prindiville SA, Reynolds HR, Dizon DS, Cooper-Arnold K, et al. Digitizing clinical trials. NPJ Digit Med 2020;3:101 [FREE Full text] [doi: 10.1038/s41746-020-0302-y] [Medline: 32821856]
25. Nazha B, Mishra M, Pentz R, Owonikoko TK. Enrollment of racial minorities in clinical trials: old problem assumes new urgency in the age of immunotherapy. Am Soc Clin Oncol Educ Book 2019 Jan;39:3-10 [FREE Full text] [doi: 10.1200/EDBK_100021] [Medline: 31099618]

Abbreviations

GAD-7: General Anxiety Disorder 7-item
IRB: institutional review board
PRO: patient-reported outcome
RCT: randomized controlled trial
SAQ: Seattle Angina Questionnaire
TAILOR-PCI: Tailored Antiplatelet Initiation to Lessen Outcomes Due to Decreased Clopidogrel Response After Percutaneous Coronary Intervention

UCSF: University of California, San Francisco

Edited by A Mavragani; submitted 05.10.21; peer-reviewed by E Olano-Espinosa, E van der Velde; comments to author 26.02.22; revised version received 04.03.22; accepted 07.03.22; published 13.06.22.

Please cite as:
Avram R, So D, Iturriaga E, Byrne J, Lennon R, Murthy V, Geller N, Goodman S, Rihal C, Rosenberg Y, Bailey K, Farkouh M, Bell M, Cagin C, Chavez I, El-Hajjar M, Ginete W, Lerman A, Levisay J, Marzo K, Nazif T, Olgin J, Pereira N
Patient Onboarding and Engagement to Build a Digital Study After Enrollment in a Clinical Trial (TAILOR-PCI Digital Study): Intervention Study
JMIR Form Res 2022;6(6):e34080
URL: https://formative.jmir.org/2022/6/e34080
doi:10.2196/34080
PMID:35699977

Original Paper

Experiences With a Postpartum mHealth Intervention During the COVID-19 Pandemic: Key Informant Interviews Among Patients, Health Care Providers, and Stakeholders

Ernani Sadural[1,2], MD, PhD; Kristen E Riley[3], PhD; Peijia Zha[2], PhD; Dula Pacquiao[2], EdD; Amanda Faust[4]

[1]Department of Obstetrics and Gynecology, Cooperman Barnabas Medical Center, RWJBarnabas Health, Livingston, NJ, United States

[2]School of Nursing, Rutgers, The State University of New Jersey, Newark, NJ, United States

[3]Graduate School of Applied and Professional Psychology, Rutgers, The State University of New Jersey, Piscataway, NJ, United States

[4]The Shaw Wellness Institute, Colgate University, Hamilton, NY, United States

Corresponding Author:
Ernani Sadural, MD, PhD
Department of Obstetrics and Gynecology
Cooperman Barnabas Medical Center
RWJBarnabas Health
94 Old Short Hills Road
Suite 3148
Livingston, NJ, 07039
United States
Phone: 1 973 885 9190
Fax: 1 973 533 4492
Email: ernani.sadural@rwjbh.org

Abstract

Background: Maternal morbidity and mortality in the United States continue to be a worsening public health crisis, with persistent racial disparities among Black women during the COVID-19 pandemic. Innovations in mobile health (mHealth) technology are being developed as a strategy to connect birthing women to their health care providers during the first 6 weeks of the postpartum period.

Objective: This study aimed to inform a process to evaluate the barriers to mHealth implementation in the context of the COVID-19 pandemic by exploring the experiences of mothers and stakeholders who were directly involved in the pilot program.

Methods: The qualitative design used GoToMeeting (GoTo) individual interviews of 13 mothers and 7 stakeholders at a suburban teaching hospital in New Jersey. Mothers were aged ≥18 years, able to read and write in English or Spanish, had a vaginal or cesarean birth at >20 weeks of estimated gestational age, and were admitted for delivery at the hospital with at least a 24-hour postpartum stay. Stakeholders were part of the hospital network's obstetrics collaborative subcommittee comprising administrators, physicians, registered nurses, and informatics. Responses were transcribed verbatim and analyzed for emerging themes. The socioecological framework provided a holistic lens for analyzing the multilevel influences on individual experiences.

Results: A total of 3 major themes were identified: mothers experienced barriers from personal situations at home and with services in the hospital and community, which were intensified by the COVID-19 pandemic; the COVID-19 pandemic negatively impacted hospital services, priorities, and individual staff; and mothers and stakeholders had positive experiences and perceptions of the mHealth intervention.

Conclusions: The use and reach of the mHealth intervention were negatively influenced by interrelated factors operating at multiple levels. The system-wide and multilevel impact of the pandemic was reflected in participants' responses, providing evidence for the need to re-evaluate mHealth implementation with more adaptable systems and structures in place using a socioecological framework.

(JMIR Form Res 2022;6(6):e37777) doi:10.2196/37777

KEYWORDS

maternal mortality; health disparity; mHealth; patient engagement; postbirth warning signs

Introduction

Background

There is a rising maternal mortality rate with a persistent racial disparity among Black mothers in the United States [1-3]. During the first year of the COVID-19 pandemic, maternal deaths continued to rise, mostly among Black and Hispanic mothers [3]. In the first 6 weeks after childbirth, the immediate postpartum period has been the deadliest time frame for birthing mothers in New Jersey [4]. In addition, Black mothers have an increased risk of postpartum readmission and associated life-threatening morbidities compared with non-Hispanic White mothers [5]. A mobile health (mHealth) intervention was instituted at a large hospital in New Jersey as a potential solution to this problem by helping mothers self-identify postbirth warning signs of postpartum complications and seek timely medical care. The program was designed to send daily SMS text messages to mothers beginning the day after discharge through the first 6 weeks postpartum. However, it is unknown which factors and to what degree they pose barriers to patient engagement. This qualitative study is part of a Type 1 hybrid effectiveness implementation pilot study, and it is limited to the qualitative portion of the project. This study is limited to the qualitative portion of the project. This exploratory study aimed to inform a process to evaluate the barriers to mHealth implementation in the context of the COVID-19 pandemic by exploring the experiences of mothers and stakeholders who were directly involved in the pilot program.

Review of the Literature

Patient education on potential postbirth warning signs has been proposed as an essential driver for strategies to eliminate delays in care seeking [6,7], particularly in the immediate postpartum period in the hospital [8]. However, the COVID-19 pandemic has often resulted in shorter postpartum stays [9,10] and fewer face-to-face professional and social interactions owing to isolation, limited visitation, and social distancing [11]. These have led to cumulative adverse effects on postpartum mental health [12,13], avoidance of the use of emergency room service [14,15], and mothers' knowledge gaps in postbirth warning signs [16]. Therefore, improving and sustaining patient education through communication and engagement channels are essential for timely care-seeking behaviors during the pandemic.

Evidence has shown postpartum mothers' willingness to use mHealth technology, particularly SMS text–based messaging, leading to opportunities for innovative new strategies for care interventions [17,18]. mHealth was found to be a feasible and acceptable intervention by mothers, especially those with diverse backgrounds, and was associated with positive benefits such as improving blood pressure monitoring and reducing racial disparities [19-21]. However, many barriers have been identified, including lack of digital access [22], low health literacy, low self-efficacy [23], perceived discrimination [24,25], and postnatal depression [26,27]. Furthermore, mHealth technology that does not factor individual patients' needs, health literacy, cultural preferences, and resources can present complex information to patients with fewer personal touchstones that might worsen health care disparities [28]. Despite the increasing use of mHealth technology, the overall rising rates of maternal mortality and racial disparities in the United States have persisted, suggesting the need to explore barriers to use among low-income mothers [29].

This qualitative study using interviews was designed to generate participants' own perceptions and descriptions of their direct experiences with the phenomenon of study. Individual interviews can elicit highly personal and subjective information that may not be evident in surveys. Participants' responses can shed light on the hidden impact of the mHealth program, allowing a holistic understanding of the phenomenon [30]. Information about barriers to the intervention, particularly those that affect the reach of at-risk Black mothers and reduce racial disparities, can be used to refine the intervention and design randomized controlled intervention trials with the mHealth intervention using the promising factors that are most likely to be effective disparity-reducing levers [31]. By doing so, the mHealth intervention can contribute to developing a more robust patient engagement platform to improve the well-being of mothers and reduce postpartum morbidity and mortality, particularly in Black mothers.

Theoretical Framework

Conceptually, the use of the mHealth intervention and its influence on health-seeking behaviors related to potential postpartum complications are based on a modified socioecological model [32,33] and the model of health services use by Anderson [34]. The model posits that behavior is integrated into a dynamic and complex network of intrapersonal factors, interpersonal processes, institutional factors, community factors, and public policy [32]. The model frames the human behaviors of seeking health information or health care services in a dynamic relationship between external influences and internal responses. The conceptual model guided the data analysis of this study.

Methods

Overview

This study was part of a pilot mHealth intervention conducted at a suburban teaching hospital in New Jersey to increase patient engagement and postpartum health service use during the COVID-19 pandemic. Resident physicians in obstetrics and gynecology (OBGYN) run the prenatal clinic under the supervision of 3 faculty attending physicians. It serves predominantly low-income, racial, and ethnic minority women from the surrounding municipalities and delivers approximately 400 babies annually. The teaching hospital is a regional perinatal center with a level III neonatal intensive care nursery and delivers approximately 5800 babies annually [35]. The intervention pilot study did not have baseline data on outcome measures to generate power size estimations. The study anticipated 125 enrollees, the expected number of deliveries for the provider group in a consecutive 3-month period. The actual sample size of 14 participants fell grossly short of our target owing to the interruption in study enrollment from the premature termination of the intervention rollout. One participant dropped

out of the mHealth intervention program and did not respond to telephone invitations for the qualitative portion of the study.

All women from the prenatal clinic and obstetrical service who delivered at the hospital and read and write in English or Spanish were eligible for enrollment in the program. The mHealth intervention used a secured SMS text–based messaging for 2-way communication to deliver a 6-item questionnaire answerable by *yes* or *no* to mothers' mobile devices. The items were designed to reinforce standard postpartum discharge education for the early identification and intervention of postbirth warning signs. Mothers who submitted any *yes* response were followed up with a phone call by a hospital provider to help determine the need for emergent care. Study participants were supposed to be sent daily SMS text messages at 9 AM from postdischarge day 1 to 6 weeks postpartum. During the initial 2-week rollout, fidelity issues were noted involving participants not receiving the daily SMS text messages and nonstudy patients incorrectly receiving the SMS text messages. The project was temporarily stopped and then restarted to reset the intervention filters and settings. Eight days later and <4 weeks from the project's go-live date, the project was halted again after discovering that the intervention did not capture 1 study participant's responses.

The failure of the intervention to send SMS text messages to some but not all participants led to the stratification of participants into 2 groups. Participants who received at least one SMS text message comprised the intervention group, and those who failed to receive any SMS text messages comprised the nonintervention group. These software bugs proved insurmountable, resulting in the termination of the project.

One participant who was reassured after reporting a potential complication through the intervention had verbalized to the principal investigator (PI) an increased sense of support generated by the SMS text–based surveys. The study team received approval from the institutional review board to support the study participants with continued monitoring for postbirth warning signs through direct weekly telephone calls to maximize patient safety. During each telephone call, the 6 questions contained in the intervention SMS text–based surveys were asked verbatim to the study participants by either the PI, coinvestigators, or the prenatal clinic physician every Monday morning until they reached their 6-week postpartum visit. Unfavorable responses to any of the 6 questions or reports of any other clinical concern were addressed immediately during the telephone encounter. The telephone encounter was documented in their electronic ambulatory medical record as a customary practice for the prenatal clinic. Further resumption of the program and the original study protocol was determined to be nonremediable. Therefore, the pilot program was prematurely terminated 3 months after the study began.

This portion of the study had a descriptive qualitative design, as described by Willis et al [36] using individual interviews. The authors developed separate interview guides for the mothers and stakeholders (Textbox 1). The results were reported according to the consolidated criteria for reporting qualitative studies.

Textbox 1. Interview guide.

Tell me about your experience with the SMS text messages from your health care provider:

Mothers

- How did it help you?
- Why was it not helpful?
- What barriers or difficulties did you experience with the text messages?
- In what way can we improve our text messages?

Stakeholders

- What was the original objective of the project?
- How has COVID affected your ability to communicate and work with patients?
- What barriers were faced with going live during COVID?
- What worked well? What has not worked well? Why do you think so?
- What lessons were learned from the process?
- What recommendations do you have for future attempts with mobile health?
- What is needed to scale up the intervention in the health system?

Ethics Approval

The institutional review boards of the hospital (19-67) and the affiliated university (Pro2020002676) approved the study protocols.

Sample

The convenience sample comprised 13 mothers who participated in the new mHealth intervention program before its termination and 7 stakeholders who were directly involved with the mHealth service. The study population included pregnant women from the hospital's prenatal clinic and obstetrical service, known as the *OB Service*, which cared for women who presented in labor

without prenatal care or an obstetrical provider on staff. Mothers were aged ≥18 years, able to read and write in English or Spanish, had a vaginal or cesarean birth at >20 weeks of estimated gestational age, and were admitted for delivery at the hospital with at least a 24-hour postpartum stay. Teen mothers were excluded from this study. The pilot study enrolled 78% (14/18) of eligible mothers, all of whom were invited to participate through in-person recruitment at the hospital clinic or as they entered the postpartum ward. Most of the mothers were ≤25 years (7/13, 54%), first-time mothers (9/13, 69%), African Americans (6/13, 46%) or Latina (6/13, 46%), Medicaid recipients (11/13, 85%), single (9/13, 69%), high school graduates (10/13, 77%), nonsmokers (12/13, 92%), and had prenatal care at the hospital (9/13, 69%).

Stakeholders were part of the hospital network's obstetrics collaborative subcommittee tasked with implementing the pilot intervention. The PI recruited the 7 key stakeholders, including key members from the prenatal clinic (medical director, nurse manager, OBGYN chief resident, project informatics leader, project executive leader, and 2 nurses from the postpartum ward).

Data Collection

The institutional review boards of the hospital and the affiliated university approved the study protocols. Study participants were recruited by telephone for a one-on-one semistructured GoToMeeting interview, a web-based, Health Insurance Portability and Accountability Act–compliant videoconference software, to gather feedback regarding the content, mode of delivery, intervention effectiveness, and barriers to provider-patient communication. One participant dropped out of the mHealth intervention program and did not respond to the telephone study invitations. The securely recorded interviews took place at home for participants (between 3 and 6 weeks postpartum) and stakeholders (within 3 months of project completion), lasted between 12 and 45 minutes, and were conducted in English for stakeholders and in Spanish or English for mothers based on their preferences. Participants were interviewed once, except for 4 mothers who required follow-up. The participants were made aware before participation that they would be provided with a US $25 digital Amazon gift card as a token of appreciation.

The PI, fluent in English and Spanish, conducted all the interviews and maintained field notes after each interview. The PI is a male OBGYN physician with prior work experience at the prenatal clinic and is employed as a corporate director by the hospital. He had no previous contact with the mothers, except in his role as a PhD student and researcher. The participants were informed that the goal of the intervention was to improve postpartum care. Mothers were assured that their responses would not affect their relationship with the hospital and clinic staff or the health services they received. They were also informed that their identity would remain confidential and would not be used to report or publish the study. All stakeholders had an existing relationship with the PI, a member of the obstetrics collaborative subcommittee overseeing the project.

Data Analysis

Interviews (including the Spanish verbatim transcripts) were transcribed verbatim into English by the PI. Interview transcripts and field notes were read independently by the first author (ES) and last author (AF) several times to gain a deep understanding of the data and develop codes for major categories of responses based on the conceptual framework model. Coded categories were entered in Microsoft Excel and discussed with the other authors to determine the barriers to and facilitators of the mHealth intervention. Several iterations of these coded responses were conducted to determine relationships in meanings across categories from which major themes emerged (Multimedia Appendix 1). Several participants provided feedback regarding the study's findings.

Rigor of the Qualitative Study

To enhance the credibility and trustworthiness of the study, reflexivity was maintained throughout the data collection. The PI analyzed field notes that chronicled nuances in participants' nonverbal responses that supported the authenticity of their verbal responses. Reflections consciously acknowledge one's values, assumptions, and goals toward the study topic; thus, the PI can clarify belief systems and subjectivities toward participants' responses [37]. Credibility and confirmability were enhanced by audiotaping interviews and verbatim transcriptions of participants' responses to capture their own descriptions of their experiences. The triangulation of findings between mothers and stakeholders enhanced the confirmation and validation of these phenomena.

Results

Overview

Multimedia Appendix 1 shows the major themes that were derived from the major categories of responses from participants. These categories were based on the socioecological framework of this study. The 3 major themes that emerged and the supporting statements from the participants are provided in the following paragraphs.

Mothers Experienced Barriers From Personal Situations at Home and With Services in the Hospital and Community, Which Were Intensified by the COVID-19 Pandemic

The pandemic has increased emotional and mental stress for mothers because of restrictive norms, resulting in social isolation and decreased social support. This was especially true for Latinas who rely on their family members for assistance. A Latina mother stated:

> *[but] I know us like Hispanics, I know the first thing we did is like we ask our parents or close relatives what to do instead of asking the doctor.*

Another Latina mother affirmed:

> *[we] get support from the child's father, the child's grandmother, aunt, niece, a whole lot of people.*

Unable to get help with transportation to the clinic, one mother said:

> *I have to take two buses. Other times, I used Uber. That was much faster.*

Mothers were overwhelmed by their own physical and emotional state, trying to cope with the competing demands between caring for the baby and other responsibilities at home, impeding their full participation in mHealth and attendance at clinic appointments. Three mothers described and shared their difficulties:

> *Postpartum depression is a lot like to deal with. It's kind of like a never-ending cycle of your mind, just never shuts up.*

> *You're so consumed with your baby that it's that might be a little overwhelming to have to answer to those questions [SMS] every single day.*

> *I'm not on my phone as much because I have moments with the baby, so my phone just dies. I don't pay attention to it.*

Mothers worried for themselves and about their babies getting COVID-19 or dying. A mother of 3 who lived with her husband, isolated from her extended family since the pandemic that took the lives of a few of their friends and family members, stated the following:

> *So, you have to be very cautious with where you're going, especially with a newborn. You have to worry. Like now, I'm afraid to go outside. We [family] communicate right now through the video, but it's still not the same.*

The pandemic has often resulted in a shorter length of postpartum stay and limited face-to-face interaction between mothers and providers, impacting relationships and communication between mothers and staff. A postpartum nurse with 20 years of experience spoke about the tremendous challenges mothers have to face after giving birth. Mentally and physically exhausted postbirth mothers are asked to assimilate more information crammed into shorter stays in the hospital, which can result in knowledge gaps. She stated the following:

> *Questions come later, and I think they come [after] maybe 2 or 3 nights at home.*

This was particularly difficult for mothers with language and literacy problems. A single Latina mother described difficulty in understanding the SMS text messages delivered in Spanish:

> *Maybe. I didn't understand [the text messages] well.*

Staffing shortages prevented the development of trust built upon the close relationship between mothers and familiar health care providers. After overhearing her nurse speak more pleasingly to the couple in the next room, a young Black mother described her experience that led to her request for a new nurse:

> *She was trying like to take over...she woke him up [partner] out of sleep just to tell him to put his mask on, she kept saying stuff. It was just annoying. It was like she was nit-picking. She was white, and she was older. She was basically trying to discriminate against us as being young parents or whatever.*

Another single Latina mother who developed severe postpartum infection and depression recalled how unsupported she felt during readmission for treatment of her infection while being separated from her newborn:

> *...and then I was in the hospital again for mastitis. It was just like I was going through a lot like mentally, personally, a lot, and it was just like a lot.*

The COVID-19 Pandemic Negatively Impacted Hospital Services, Priorities, and Individual Staff

Mothers and stakeholders expressed frustration with technical issues and the lack of resources for troubleshooting problems. The project's technical team spoke of difficulty keeping the project moving:

> *...the biggest barrier was that some of the I.T. team and other people involved with the project were redeployed into other areas where they were needed to address the needs and the competing priorities of COVID.*

Several stakeholders pointed out the compounding circumstances resulting from the change in ownership of the mHealth vendor:

> *One of the biggest challenges...was people that we had been working with, all along, left the company, either by choice or by design.*

The OBGYN chief resident charged with the operational rollout of the mHealth intervention emphasized the following:

> *I think we have less resources. People are focusing on other areas. I feel like COVID kind of wiped that out.*

Frontline nurses complained about staffing shortages, increased workload, lack of time with patients, frequent reassignment, and burnout during the COVID-19 pandemic. When asked how her nursing colleagues felt during the project, nurse stakeholders described the following:

> *...it's not just our hospital, and I think it's everywhere, the help is not there.*

Many nursing administrators were reassigned to the bedside, often in critical care units full of patients with COVID-19. Stakeholders spoke of their underlying fear and experience of COVID-19 infection:

> *Everyone's still kind of scared, everyone's on edge.*

A physician stakeholder said:

> *Nurses are burnt out from this year. They had a year of doing COVID.*

Stakeholders identified the need for greater collaboration with community agencies. Some (2/13, 15%) mothers found support from a regional maternal-child health consortium program, but most (7/13, 54%) mothers turned to hospital-based or social media–based virtual communities for support. A postpartum nurse observed the following:

> *They sometimes go to social media for that, they go on, you know, on Facebook or Instagram, and ask their friends or our mom's support group, "what, is this normal?"*

Mothers and Stakeholders had Positive Experiences and Perceptions of the mHealth Intervention

Mothers felt relieved of their anxieties in caring for their newborn by being able to reach out to their care providers from home to seek information, reinforce their knowledge, and address problems, as stated by some mothers:

> *I use them to remind myself like what I should be looking for, or any other problems come up. as well as reassure myself and know that y'all are worried about us and how we do in and stuff.*
>
> *So that kinda calms me down that everything is fine...that was a benefit definitely for patients that have anxiety issues.*
>
> *In the beginning, probably the information for me was fresh, and I remember all the information that I obtained...because you don't remember every little thing.*
>
> *You're not worried about yourself, you're so worried about making sure your baby is ok.*

The intervention was instrumental in building trusting relationships between mothers and providers. A single Latina mother who experienced postpartum depression and was readmitted with an infection said the following:

> *I just felt a little bit more comfortable coming to this hospital, even if it was like with whatever Doctor was on call. You had trust in them.*

Another Latina mother who had complications during her 2 prior pregnancies stated the following:

> *M.B. still communicates with me, which is good.*

A physician stakeholder noted the following:

> *I think those are things that make people connect better if they know who their team is and even knowing the residents.*

There was consensus among stakeholders that the intervention improved their ability to provide care for disadvantaged populations. Two stakeholders commented the following:

> *More and more people need help in many ways, with a lot of things. These patients deserve the best care.*
>
> *I try to pause whatever's going on and try to commonly address the hesitancies that I can read and asking how maybe I can make them more comfortable, or is there anything else they'd like to address. And then, if it's a cultural difference, figuring out how to recognize that, how to respect that, and kind of exploring with them what else they need from me as a provider.*
>
> *It was good for me because she was able to point out exactly like what she was talking about because I was actually showing her.*

Two stakeholders emphasized that mHealth facilitated their ability to monitor patients closely:

> *It was a great way to keep a pulse on them, making sure they're doing OK.*
>
> *And also, if you're reaching out, you're going to pick up stuff.*

Two other stakeholders recognized mHealth as a pathway to quality improvement:

> *...taking not just an emergency contact, but a second contact, as well...partnering...with the community in finding different clinics and resources.*
>
> *...some of our residents are very social service-oriented. Currently, our residency group is super, super helpful in connecting resources, and we've been partnering with neighboring hospitals with specialties that can accommodate charity care.*

The project executive aptly summarized how mHealth improved maternal and childcare in the hospital:

> *I think that with technology, we're able to really bring care to where the patient lives without some of the confounding factors that may prevent that engagement otherwise, so I think there's a huge opportunity to improve outcomes and reduce disparities and provide equity in health care with the use of technology. Prior to the pandemic, we've had difficulty engaging Black Moms in the postpartum period...through our virtual support groups, we have seen an increase and increased engagement, where the barriers of transportation and childcare and timing have essentially been removed.*

Discussion

Principal Findings

The findings of this study revealed that the experiences of mothers and stakeholders with the mHealth intervention were largely influenced by socioecological factors on the individual mother; her family; and the social support network stemming from the shifting demands of the hospital organization, service delivery, and the community. Mothers who were uninsured and those with limited resources at home (social support) and in the community (transportation) experienced greater vulnerability because of COVID-19.

Although mHealth allowed greater access to providers, this was not uniformly experienced by mothers with limited resources, reflected in the amenities provided by their cell phone contracts and access to Wi-Fi services. Difficulties described by mothers, ranging from technical barriers to receiving SMS text messages to understanding SMS text message content, highlight the need to consider the population's sociodemographic characteristics and digital health literacy to achieve simplicity in intervention use. The development of digital health literacy and language assistance are essential for effective electronic communication with providers in the new digital-based context of care delivery [38]. Face-to-face assessment of mothers and providing access to a phone receptionist who can facilitate personal assistance when needed were examples raised by mothers and stakeholders to help mothers who might lack the competencies to navigate the digital health care system. An integrative approach to the implementation of digital interventions should be built into the service to ensure accessibility and inclusivity [39].

The lack of continuity of care from hospital discharge to postpartum care hampered relationships and communication between mothers and staff. There is a greater risk of perceived unfair treatment and discrimination by patients when there is a lack of prior relationship and trust built with their care providers. Staff development in cross-cultural communication emphasizing respect, sensitivity, and compassion should be conducted to foster a nurturing environment for implementing the intervention [40]. Cross-cultural champions and mentors can offer consistent guidance and assistance in conflict management.

The motivation for the project's inception was to mitigate the impact of the pandemic on health care delivery and follow-up postpartum care for mothers and their babies. However, the unprecedented impact of COVID-19 on hospital organization and service delivery, eventually shifting its priorities to pressing demands, caused the early termination of the project. The findings of this study validate the cascading effects of a global or societal event on organizations, including health care systems, the community, and their residents. Participants' experiences provided evidence of the significant role of using an ecological framework in understanding the microlevel experiences of mothers and hospital stakeholders with postpartum care. Including several authors, some of whom have expertise in qualitative research, enhanced the interpretive consensus and reliability of the findings. The framework enabled a holistic perspective of the phenomenon of an mHealth intervention in postpartum follow-up care.

The pandemic has created many barriers to the implementation, making it difficult to fully assess the outcomes of the intervention. Nevertheless, participants identified some advantages and potential of the program, particularly the reassuring and calming effects of the SMS text messages. This was important in the face of increased psychological distress and anxiety for mothers in the age of COVID-19 and the need for timely interventions to mitigate mental disorders [13,41]. The findings of the study affirm the need to examine health care services within a socioecological framework and to plan interventions using the same framework to enhance preparedness for multilevel influences and impact on outcomes.

Lessons Learned for Implementation

Barriers to implementation may arise at the patient, provider team, organizational, market, or policy level of health care delivery [42], requiring a multilevel change approach. Our implementation efforts were deeply intertwined within the context of the global COVID-19 pandemic. COVID-19 had a negative impact on implementing the intervention at the organizational level owing to the overarching focus on, and redeployment of, clinical informatics resources toward the care of patients with COVID-19. Specifically, all stakeholders from the hospital spoke of the compounding disruptive effects of COVID-19 layered on the loss of key personnel at the vendor company. The project was halted several times during the global pandemic because of the redeployment of staff and resources at both the hospital and vendor company. Combined with the challenge of vendor ownership transition, the disruption of hospital operations because of the pandemic was the key driver of implementation failure.

During the initial intervention launch, stakeholders expressed frustration regarding the lack of vendor support in resolving software bugs. In particular, the informatics project leader spoke of the loss of historical knowledge transfer from the original vendor team to the new team as the key barrier to resolving the implementation issues. This assessment is supported in the literature, in that operational capabilities such as the degree of agility, quality of resources, and quality of cooperation can be more significant than technological capabilities for organizational e-readiness for digital transformation [43].

Frontline providers in this study spoke about mental and physical exhaustion and feeling burnt out after a year of facing COVID-19. This is consistent with studies reporting significant hospital operational challenges brought on by the COVID-19 pandemic, including reduced bed capacity, shortages of health care personnel, supply chain disruption, and burnout among health care providers [44,45]. A cadre of consistent nurses, residents, and other health care personnel should be assured once the pandemic's impact is stabilized.

Successful digital technology implementation requires organizational readiness for change with a sustainable organizational leadership commitment to ensure the allocation of a dedicated budget and personnel for the project [43,46]. On the basis of the feedback from stakeholders in this study, the possibility of a larger implementation of mHealth in other hospital services is more likely to occur, requiring sustained organizational commitment, e-readiness, and support. The significant influence of operational capabilities rather than technological capabilities has been suggested as a key feature of organizational e-readiness [43]. Unfortunately, the assessment of organizational e-readiness before implementation has proved elusive during the pandemic.

Engaging the broader community is important for health promotion and addressing gaps in care. Organizational leadership involvement in developing partnerships and collaborating with other agencies and the community can generate more comprehensive support for patients and their families in the community. Services such as transportation, pharmacy, police, and food vendors are essential for ensuring comprehensive health care services.

Integral to quality improvement is how outcomes are measured and correlated with the structure and processes of health care [47]. The results of this study and other quantitative measures of outcomes should be integrated into future mHealth designs and quality improvement.

Limitations

There were significant limitations to this study. It would be difficult to know what would have happened without the disruptive impact of a global pandemic. The extremely limited sample size and technical difficulties with the delivery and routing of SMS text messages precluded deeper insights into implementation barriers and facilitators. However, the shifting socioecological context and confounders of the COVID-19 pandemic were central to the diffusion, dissemination, implementation, and ultimately, failure of the mHealth

intervention, leaving us valuable lessons for future implementations.

Although most of our study sample consisted of Black and Latina mothers, the small sample size, which was only one-tenth of the planned convenience sample, introduced considerable selection bias, diminishing internal and external validity. This may have been exacerbated by the addition of a token bias through economic incentives, as the fear of losing the incentive could have influenced the participants' answers. Our stakeholder sample might also exhibit selection bias, as stakeholders might have had higher intrinsic motivation and interest in the research topic than the nonparticipants.

The role of the PI as a clinician, hospital administrator, and researcher may have fostered social desirability bias by inhibiting open discussion with study patients who may have believed that their responses might affect their treatment. The PI played a significant role in the implementation process, resulting in frequent contact with study participants and stakeholders, which may have led to researcher confirmation bias. Data saturation could not be ensured given the small study sample and the fact that one of the original study participants did not respond to requests to be interviewed.

Further research is needed to pilot test the intervention, preferably using a multiphase optimization strategy to test various intervention strategies identified from the study. The multiphase optimization strategy design offers an easy way to assess the feasibility of components before testing them with a fully powered experiment [48], and it can identify potentially challenging combinations of remotely delivered intervention components. Once the SMS text messaging intervention has been fine-tuned by this approach, a specific, measurable, achievable, relevant, and time-bound adaptive intervention study could determine which components of the larger intervention work well (eg, implicit bias training, doulas, patient navigators, and travel vouchers), for whom, in what combination, and when.

Conclusions

The socioecological framework enhanced the understanding of the experiences of mothers and stakeholders within the context of the COVID-19 pandemic and its impact on hospitals, communities, and families. The system-wide and multilevel impact of the pandemic was reflected in participants' responses, providing evidence for the need to re-evaluate mHealth implementation with more adaptable systems and structures in place. The ecological framework helped to analyze why individual-level plans may not work in changing social contexts. Planned interventions must go beyond the microsystem level to prepare organizations and communities to support individual-level interventions. Addressing the social determinants of health in the community and organizational system capacity are critical to achieving effective care outcomes at the individual level.

Acknowledgments

This study was supported by a grant from the Harvey E Nussbaum Foundation. The authors would like to thank Liliana Morales, MPH, RN; Chelsey Russ, MD; and Zaire Ali, MPH, for their support throughout the research and writing of this paper.

Conflicts of Interest

In accordance with ethical obligations as a researcher, the principal investigator (ES) is reporting a financial interest as an employee of RWJBarnabas Health, a health care organization that may be affected by the research reported in this study. This has been fully disclosed to the institutional review board at Rutgers, The State University of New Jersey. An approved plan for managing potential conflicts arising from this involvement is in place.

Multimedia Appendix 1
Data analysis matrix using the socioecological framework.
[DOCX File , 16 KB - formative_v6i6e37777_app1.docx]

References

1. GBD 2015 Maternal Mortality Collaborators. Global, regional, and national levels of maternal mortality, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet 2016 Oct 08;388(10053):1775-1812 [FREE Full text] [doi: 10.1016/S0140-6736(16)31470-2] [Medline: 27733286]
2. Petersen EE, Davis NL, Goodman D, Cox S, Mayes N, Johnston E, et al. Vital signs: pregnancy-related deaths, United States, 2011-2015, and strategies for prevention, 13 states, 2013-2017. MMWR Morb Mortal Wkly Rep 2019 May 10;68(18):423-429 [FREE Full text] [doi: 10.15585/mmwr.mm6818e1] [Medline: 31071074]
3. Maternal Mortality Rates in the United States, 2020. Centers for Disease Control and Prevention. 2022. URL: https://www.cdc.gov/nchs/data/hestat/maternal-mortality/2020/maternal-mortality-rates-2020.htm [accessed 2022-03-01]
4. Trends in Statewide Maternal Mortality: 2009-2013. New Jersey Department of Health. 2016. URL: https://nj.gov/health/fhs/maternalchild/documents/nj_maternal_mortality_trends_2009_2013.pdf [accessed 2021-04-01]
5. Aziz A, Gyamfi-Bannerman C, Siddiq Z, Wright JD, Goffman D, Sheen JJ, et al. Maternal outcomes by race during postpartum readmissions. Am J Obstet Gynecol 2019 May;220(5):484.e1-484.10 [FREE Full text] [doi: 10.1016/j.ajog.2019.02.016] [Medline: 30786255]

6. Mulugeta AK, Giru BW, Berhanu B, Demelew TM. Knowledge about birth preparedness and complication readiness and associated factors among primigravida women in Addis Ababa governmental health facilities, Addis Ababa, Ethiopia, 2015. Reprod Health 2020 Jan 29;17(1):15 [FREE Full text] [doi: 10.1186/s12978-020-0861-z] [Medline: 31996221]
7. Hear Her Campaign. Centers for Disease Control and Prevention. 2022. URL: https://www.cdc.gov/hearher/ [accessed 2022-03-01]
8. Suplee PD, Kleppel L, Santa-Donato A, Bingham D. Improving postpartum education about warning signs of maternal morbidity and mortality. Nurs Womens Health 2017 Dec;20(6):552-567. [doi: 10.1016/j.nwh.2016.10.009] [Medline: 27938796]
9. Bornstein E, Gulersen M, Husk G, Grunebaum A, Blitz MJ, Rafael TJ, et al. Early postpartum discharge during the COVID-19 pandemic. J Perinat Med 2020 Nov 26;48(9):1008-1012 [FREE Full text] [doi: 10.1515/jpm-2020-0337] [Medline: 32845868]
10. Rochelson B, Nimaroff M, Combs A, Schwartz B, Meirowitz N, Vohra N, et al. The care of pregnant women during the COVID-19 pandemic - response of a large health system in metropolitan New York. J Perinat Med 2020 Jun 25;48(5):453-461 [FREE Full text] [doi: 10.1515/jpm-2020-0175] [Medline: 32432568]
11. How do pregnant people or new mothers stay safe? What rules are in place for labor and delivery during COVID-19? New Jersey Department of Health. 2021. URL: https://covid19.nj.gov/faqs/nj-information/slowing-the-spread/how-do-pregnant-people-or-new-mothers-stay-safe-what-rules-are-in-place-for-labor-and-delivery-during-covid-19 [accessed 2022-03-01]
12. Jackson L, De Pascalis L, Harrold JA, Fallon V, Silverio SA. Postpartum women's psychological experiences during the COVID-19 pandemic: a modified recurrent cross-sectional thematic analysis. BMC Pregnancy Childbirth 2021 Sep 17;21(1):625 [FREE Full text] [doi: 10.1186/s12884-021-04071-2] [Medline: 34530772]
13. Yan H, Ding Y, Guo W. Mental health of pregnant and postpartum women during the coronavirus disease 2019 pandemic: a systematic review and meta-analysis. Front Psychol 2020 Nov 25;11:617001 [FREE Full text] [doi: 10.3389/fpsyg.2020.617001] [Medline: 33324308]
14. Castillo EM, Cronin AO, Vilke GM, Killeen JP, Brennan JJ. 169 emergency department utilization trends during the COVID-19. Ann Emerg Med 2020 Oct;76(4):S66. [doi: 10.1016/j.annemergmed.2020.09.181]
15. Gutovitz S, Pangia J, Finer A, Rymer K, Johnson D. Emergency department utilization and patient outcomes during the COVID-19 pandemic in America. J Emerg Med 2021 Jun;60(6):798-806 [FREE Full text] [doi: 10.1016/j.jemermed.2021.01.002] [Medline: 33581990]
16. Adams YJ, Young J. Perceptions of postpartum teaching and knowledge of warning signs among Black mothers. West J Nurs Res 2022 Jan;44(1):31-41. [doi: 10.1177/01939459211043939] [Medline: 34666580]
17. Logsdon MC, Lauf A, Stikes R, Revels A, Vickers-Smith R. Partnering with new mothers to develop a smart phone app to prevent maternal mortality after hospital discharge: a pilot study. J Adv Nurs 2020 Jan;76(1):324-327. [doi: 10.1111/jan.14219] [Medline: 31588581]
18. Shorey S, Lau YY, Dennis CL, Chan YS, Tam WW, Chan YH. A randomized-controlled trial to examine the effectiveness of the 'Home-but not Alone' mobile-health application educational programme on parental outcomes. J Adv Nurs 2017 Sep;73(9):2103-2117. [doi: 10.1111/jan.13293] [Medline: 28276086]
19. Hirshberg A, Downes K, Srinivas S. Comparing standard office-based follow-up with text-based remote monitoring in the management of postpartum hypertension: a randomised clinical trial. BMJ Qual Saf 2018 Nov;27(11):871-877. [doi: 10.1136/bmjqs-2018-007837] [Medline: 29703800]
20. Dol J, Aston M, McMillan D, Tomblin Murphy G, Campbell-Yeo M. Effectiveness of a postpartum text message program (essential coaching for every mother) on maternal psychosocial outcomes: protocol for a randomized controlled trial. JMIR Res Protoc 2021 Mar 25;10(3):e27138 [FREE Full text] [doi: 10.2196/27138] [Medline: 33764309]
21. Hirshberg A, Sammel MD, Srinivas SK. Text message remote monitoring reduced racial disparities in postpartum blood pressure ascertainment. Am J Obstet Gynecol 2019 Sep;221(3):283-285. [doi: 10.1016/j.ajog.2019.05.011] [Medline: 31121137]
22. Eruchalu CN, Pichardo MS, Bharadwaj M, Rodriguez CB, Rodriguez JA, Bergmark RW, et al. The expanding digital divide: digital health access inequities during the COVID-19 pandemic in New York city. J Urban Health 2021 Apr;98(2):183-186 [FREE Full text] [doi: 10.1007/s11524-020-00508-9] [Medline: 33471281]
23. Chung K, Cho HY, Kim YR, Jhung K, Koo HS, Park JY. Medical help-seeking strategies for perinatal women with obstetric and mental health problems and changes in medical decision making based on online health information: path analysis. J Med Internet Res 2020 Mar 04;22(3):e14095 [FREE Full text] [doi: 10.2196/14095] [Medline: 32130139]
24. Vedam S, Stoll K, Taiwo TK, Rubashkin N, Cheyney M, Strauss N, GVtM-US Steering Council. The Giving Voice to Mothers study: inequity and mistreatment during pregnancy and childbirth in the United States. Reprod Health 2019 Jun 11;16(1):77 [FREE Full text] [doi: 10.1186/s12978-019-0729-2] [Medline: 31182118]
25. Crear-Perry J, Correa-de-Araujo R, Lewis Johnson T, McLemore MR, Neilson E, Wallace M. Social and structural determinants of health inequities in maternal health. J Womens Health (Larchmt) 2021 Feb;30(2):230-235 [FREE Full text] [doi: 10.1089/jwh.2020.8882] [Medline: 33181043]

26. Dol J, Richardson B, Murphy GT, Aston M, McMillan D, Campbell-Yeo M. Impact of mobile health interventions during the perinatal period on maternal psychosocial outcomes: a systematic review. JBI Evid Synth 2020 Jan;18(1):30-55. [doi: 10.11124/JBISRIR-D-19-00191] [Medline: 31972680]
27. Singh K, Drouin K, Newmark LP, Rozenblum R, Lee J, Landman A, et al. Developing a framework for evaluating the patient engagement, quality, and safety of mobile health applications. Issue Brief (Commonw Fund) 2016 Feb;5:1-11. [Medline: 26934758]
28. Sentell T, Foss-Durant A, Patil U, Taira D, Paasche-Orlow MK, Trinacty CM. Organizational health literacy: opportunities for patient-centered care in the wake of COVID-19. Qual Manag Health Care 2021 Jan 1;30(1):49-60. [doi: 10.1097/QMH.0000000000000279] [Medline: 33229999]
29. Neggers YH. Trends in maternal mortality in the United States. Reprod Toxicol 2016 Sep;64:72-76. [doi: 10.1016/j.reprotox.2016.04.001] [Medline: 27063184]
30. Denzin NK, Lincoln YS. The SAGE Handbook of Qualitative Research. 4th edition. Thousand Oaks, CA, USA: Sage Publications; 2011.
31. Lee CS, Larson DB. Beginner's guide to practice quality improvement using the model for improvement. J Am Coll Radiol 2014 Dec;11(12 Pt A):1131-1136. [doi: 10.1016/j.jacr.2014.08.033] [Medline: 25467725]
32. McLeroy KR, Bibeau D, Steckler A, Glanz K. An ecological perspective on health promotion programs. Health Educ Q 1988;15(4):351-377. [doi: 10.1177/109019818801500401] [Medline: 3068205]
33. Sword W. A socio-ecological approach to understanding barriers to prenatal care for women of low income. J Adv Nurs 1999 May;29(5):1170-1177. [doi: 10.1046/j.1365-2648.1999.00986.x] [Medline: 10320501]
34. Andersen RM. Revisiting the behavioral model and access to medical care: does it matter? J Health Soc Behav 1995 Mar;36(1):1-10. [Medline: 7738325]
35. New Jersey Maternal Data Center: 2018 Maternal Health Hospital Report Cards. New Jersey Department of Health. 2020. URL: https://nj.gov/health/maternal/morbidity/mhh_reportcard/2018/all_hospitals.shtml?name=Saint_Barnabas_Medical_Center [accessed 2022-03-01]
36. Willis DG, Sullivan-Bolyai S, Knafl K, Cohen MZ. Distinguishing features and similarities between descriptive phenomenological and qualitative description research. West J Nurs Res 2016 Sep;38(9):1185-1204. [doi: 10.1177/0193945916645499] [Medline: 27106878]
37. Denzin NK, Lincoln YS. The Landscape of Qualitative Research. 4th edition. Thousand Oaks, CA, USA: Sage Publications; 2013.
38. Palumbo R, Nicola C, Adinolfi P. Addressing health literacy in the digital domain: insights from a literature review. Kybernetes 2022;51(13):82-97. [doi: 10.1108/k-07-2021-0547]
39. Watson A, Wilkinson TM. Digital healthcare in COPD management: a narrative review on the advantages, pitfalls, and need for further research. Ther Adv Respir Dis 2022;16:17534666221075493 [FREE Full text] [doi: 10.1177/17534666221075493] [Medline: 35234090]
40. Eubanks RL, McFarland MR, Mixer SJ, Muñoz C, Pacquiao DF, Wenger AF. Chapter 4: cross-cultural communication. J Transcult Nurs 2010 Oct 28;21(4_suppl):137S-150S. [doi: 10.1177/1043659610374322]
41. Wang PS, Angermeyer M, Borges G, Bruffaerts R, Tat Chiu W, DE Girolamo G, et al. Delay and failure in treatment seeking after first onset of mental disorders in the World Health Organization's World Mental Health Survey Initiative. World Psychiatry 2007 Oct;6(3):177-185 [FREE Full text] [Medline: 18188443]
42. Ferlie EB, Shortell SM. Improving the quality of health care in the United Kingdom and the United States: a framework for change. Milbank Q 2001;79(2):281-315 [FREE Full text] [doi: 10.1111/1468-0009.00206] [Medline: 11439467]
43. Kruszyńska-Fischbach A, Sysko-Romańczuk S, Rafalik M, Walczak R, Kludacz-Alessandri M. Organizational e-readiness for the digital transformation of primary healthcare providers during the COVID-19 pandemic in Poland. J Clin Med 2021 Dec 27;11(1):133 [FREE Full text] [doi: 10.3390/jcm11010133] [Medline: 35011873]
44. Sultana A, Sharma R, Hossain MM, Bhattacharya S, Purohit N. Burnout among healthcare providers during COVID-19: challenges and evidence-based interventions. Indian J Med Ethics 2020;V(4):1-6 [FREE Full text] [doi: 10.20529/IJME.2020.73] [Medline: 34018959]
45. Wu H, Soe MM, Konnor R, Dantes R, Haass K, Dudeck MA, National Healthcare Safety Network. Hospital capacities and shortages of healthcare resources among US hospitals during the coronavirus disease 2019 (COVID-19) pandemic, National Healthcare Safety Network (NHSN), March 27-July 14, 2020. Infect Control Hosp Epidemiol (forthcoming) 2021 Jun 24:1-4 [FREE Full text] [doi: 10.1017/ice.2021.280] [Medline: 34167599]
46. Fowe IE. Evaluating organizational readiness for change in the implementation of telehealth and mobile health interventions for chronic disease management. AMIA Jt Summits Transl Sci Proc 2021 May 17;2021:210-219 [FREE Full text] [Medline: 34457135]
47. Donabedian A. Evaluating the quality of medical care. 1966. Milbank Q 2005;83(4):691-729 [FREE Full text] [doi: 10.1111/j.1468-0009.2005.00397.x] [Medline: 16279964]
48. Kugler KC, Balantekin KN, Birch LL, Savage JS. Application of the multiphase optimization strategy to a pilot study: an empirical example targeting obesity among children of low-income mothers. BMC Public Health 2016 Nov 22;16(1):1181 [FREE Full text] [doi: 10.1186/s12889-016-3850-y] [Medline: 27876027]

Abbreviations

mHealth: mobile health
OBGYN: obstetrics and gynecology
PI: principal investigator

Edited by A Mavragani; submitted 07.03.22; peer-reviewed by J Dol, B Nievas Soriano; comments to author 26.03.22; revised version received 15.04.22; accepted 22.05.22; published 13.06.22.

Please cite as:
Sadural E, Riley KE, Zha P, Pacquiao D, Faust A
Experiences With a Postpartum mHealth Intervention During the COVID-19 Pandemic: Key Informant Interviews Among Patients, Health Care Providers, and Stakeholders
JMIR Form Res 2022;6(6):e37777
URL: https://formative.jmir.org/2022/6/e37777
doi:10.2196/37777
PMID:35699998

Original Paper

Geosocial Networking App Use Associated With Sexual Risk Behavior and Pre-exposure Prophylaxis Use Among Gay, Bisexual, and Other Men Who Have Sex With Men: Cross-sectional Web-Based Survey

Laurel P Gibson[1], MA; Emily B Kramer[1], BA; Angela D Bryan[1], PhD

Department of Psychology and Neuroscience, University of Colorado Boulder, Boulder, CO, United States

Corresponding Author:
Laurel P Gibson, MA
Department of Psychology and Neuroscience
University of Colorado Boulder
Muenzinger D244
345 UCB
Boulder, CO, 80309
United States
Phone: 1 303 492 8662
Fax: 1 303 492 2967
Email: laurel.gibson@colorado.edu

Abstract

Background: In the United States, geosocial networking (GSN) apps (ie, mobile dating apps) have become central to dating and sexual interactions in recent years. Among gay, bisexual, and other men who have sex with men (GBM), these apps play an important role in reducing barriers and facilitating partner seeking. However, despite these benefits, there are concerns that these apps may facilitate risky sexual behavior and transmission of sexually transmitted infections (STIs) among GBM.

Objective: This study aimed to examine the association between GSN app use and sexual risk in a US sample of GBM.

Methods: Using a cross-sectional design, respondents (N=223) completed a web-based survey assessing their use of GSN apps, sexual risk and protective behaviors, HIV serostatus, and previous STI diagnoses.

Results: Respondents were aged 21-78 (mean 31.90, SD 10.06) years and 69.5% (155/223) were non-Hispanic White. The sample included respondents from 40 states and the District of Columbia. Nearly half (104/223, 47%) of the participants reported using GSN apps. GSN users were more likely to report past-year condomless anal intercourse (P<.001), 3 or more sexual partners in the previous year (P<.001), and a previous STI diagnosis (P=.001) than nonusers. GSN users also reported more frequent use of recreational drugs before sex (P=.001), alcohol use before sex (P<.001), and cannabis use before sex (P=.01). Interestingly, GSN users were also more likely to report having ever taken an HIV test (P<.001) and using pre-exposure prophylaxis (P=.03). The rates of HIV seropositivity did not differ significantly between GSN users and nonusers (P=.53). Among the subset of GSN users, 38 participants reported using only GBM-specific GSN apps (eg, Grindr), whereas 27 participants reported using only sexuality nonspecific GSN apps (eg, Tinder). Exclusive users of GBM–specific apps reported more frequent recreational drug use before sex (P=.01) and were also more likely to report past-year condomless anal intercourse (P<.001), 3 or more sexual partners in the previous year (P=.004), a previous STI diagnosis (P=.002), and HIV testing (P=.003). Alcohol use before sex, cannabis use before sex, pre-exposure prophylaxis use, and HIV rates were similar between both groups (P>.11).

Conclusions: The findings suggest that GSN apps may be a useful pathway for interventions aimed at reducing STI risk in GBM. Future prospective studies should examine how risk levels change after the initiation of GSN app use.

(JMIR Form Res 2022;6(6):e35548) doi:10.2196/35548

KEYWORDS

dating app; mobile dating; hookup; gay; pre-exposure prophylaxis; sexual risk; HIV; STI; mobile phone

Introduction

Background

Although gay, bisexual, and other men who have sex with men (GBM) make up a small fraction of the US population, they are disproportionately affected by HIV, comprising nearly 70% of new HIV cases in the United States in 2018 [1]. Furthermore, GBM experience elevated rates of several sexually transmitted infections (STIs) beyond HIV, including syphilis, human papillomavirus, and hepatitis [2-5]. The disproportionate burden of STIs experienced by this community may be driven, in part, by higher rates of risky sexual behaviors, which are known to increase susceptibility to STIs, including HIV [6].

The US Centers for Disease Control and Prevention have identified GBM as a primary target group for HIV and STI prevention efforts, emphasizing the need for tailored interventions aimed at reducing risk behaviors among GBM. As such, there is a critical need to identify factors associated with sexual risk among GBM in order to inform our approach to mitigating the transmission of STIs, including HIV.

The use of geosocial networking (GSN) apps (ie, mobile dating apps) may be a key factor associated with sexual risk in GBM. GSN apps are smartphone apps that use GPS data to allow individuals seeking romantic or sexual partners to quickly connect and chat with other users nearby. The popularity of GSN apps (eg, Tinder and Grindr) has surged in recent years [7], most notably during the COVID-19 pandemic [8], and there is broad consensus that these apps have become central to dating and sexual interactions, particularly among GBM. According to recent survey data, 55% of lesbian, gay, and bisexual (LGB) adults, including GBM, have used a dating website or app at some point compared with 28% of their heterosexual counterparts [9]. The popularity of dating apps among LGB adults has been attributed to their role in addressing the barriers to meeting potential sexual partners typically faced by members of sexual minority groups, such as stigma and a desire for anonymity and discretion among those who prefer not to publicly identify as LGB [10].

Despite these benefits, there is also evidence supporting concerns that GSN apps may facilitate risky sexual behavior and transmission of STIs among GBM. Although it is possible that GSN apps may directly promote riskier behaviors, it is also possible that the innovative features of these apps that facilitate connections among LGB individuals attract individuals seeking to engage in risky behaviors (ie, self-selection) [11]. For instance, the use of GSN apps may reflect greater engagement in subcultures with more accepting norms toward sexual risk taking (eg, sexualized drug use) [12,13]. Regardless of the mechanism, the rates of STIs are high among app-using GBM [6], and a recent meta-analysis examining GSN app use among GBM identified a greater likelihood of STI (ie, syphilis, gonorrhea, and chlamydia) diagnoses among app-using GBM than among nonusers [14]. Counterintuitively, the same meta-analysis found no significant differences in HIV infection rates between app users and nonusers [14], suggesting that there may be some nuance to the association between app use and sexual risk among GBM.

Notably, high rates of some risky sexual behaviors—such as condomless anal intercourse (CAI), group sex, and sexualized drug use—have been reported among app-using GBM [15-19], but reviews have highlighted the need to directly compare users and nonusers [6,14]. A few recent studies have documented a higher likelihood of CAI, group sex, and a greater number of concurrent sexual partnerships among app-using GBM than among nonusers [15,20-24]. However, most studies comparing the risk behaviors of app-using and nonusing GBM have relied on homogenous samples (eg, GBM from Washington, District of Columbia [17]) and are from non-US countries (eg, China [22-24] and Nigeria [21]), making it unclear whether their findings generalize to more diverse samples of GBM from across the United States.

Despite a number of studies showing higher risk behavior in app-using GBM, there are also data suggesting that app-using GBM may also engage in higher levels of *protective* behaviors than nonusers. Several studies have found that app users report being tested for HIV more frequently than their app nonusing counterparts [17,24,25], and one study of GBM receiving HIV and STI screening in San Diego found that GBM who reported using Grindr were more likely to be regular pre-exposure prophylaxis (PrEP) users than those who did not report using Grindr [26]. However, it is unclear whether this finding generalizes to users of GSN apps beyond Grindr, and more studies on the association between GSN app use and PrEP use are needed. Taken together, these findings paint a complex picture, suggesting that although app users have been shown to engage in some risky behaviors more frequently than nonusers, they may also be more likely to engage in risk mitigation and prevention practices, perhaps recognizing their risk of STI infection.

It is important to note that most studies examining GSN app use among GBM recruited samples *directly from dating apps* (eg, [15-19]), which underscores the need for additional research comparing risky sexual behaviors and relevant health outcomes between app-using GBM and their app nonusing counterparts. Furthermore, there is a dearth of research examining differences in risk behaviors among GBM primarily using GSN apps specific to lesbian, gay, bisexual, transgender, and queer/questioning individuals (eg, Grindr) and GBM primarily using sexuality nonspecific apps (eg, Tinder). As sexual partnerships pursued online differ from those pursued through in-person venues [14], it is also possible that risky behaviors may vary by *type* of web-based venue. Supporting the self-selection argument, research suggests that GBM-specific apps are more often used for *hookups* with casual partners than for pursuing long-term relationships [27].

This Study

In summary, further research is needed to understand how risk profiles differ between app-using and app nonusing GBM, as well as to identify the types of GSN apps that may be associated with higher levels of risk among this population. Thus, this study aimed to address these gaps in the literature by examining sexual risk and protective behaviors as well as HIV and STI prevalence among a diverse sample of GBM including users of GSN apps specific to GBM, users of sexuality nonspecific apps,

and app nonusers. We hypothesized that GBM who use GSN apps would report greater HIV and STI prevalence, as well as higher levels of sexual risk behaviors, compared to GBM who do not use GSN apps. In addition, we hypothesized that GBM who use GSN apps would report greater *protective* behaviors, such as pre-PrEP use and HIV testing, despite or perhaps *because of* heightened levels of sexual risk behavior. Finally, we conducted exploratory analyses to examine whether these outcomes differ between GBM using *only* GBM-specific GSN apps and GBM using *only* sexuality nonspecific GSN apps. Given the increasing popularity of these apps [7,8] as well as the increasing rates of STIs in the United States [4], this is both a necessary and timely avenue of research to pursue.

Methods

Recruitment

This study used baseline data from a larger experimental study aimed at understanding the causal impact of discrimination on various health behaviors. Respondents were recruited in November 2020 from the web-based crowdsourcing platform Prolific and were eligible to participate if they (1) identified as gay, bisexual, or queer; (2) identified as cisgender male; and (3) were US residents. Prolific is similar to Amazon's Mechanical Turk; however, it is geared toward academic research, allows niche recruitment, and offers higher data quality [28]. Eligible respondents were invited to complete an anonymous survey assessing their engagement in various health behaviors. Respondents provided informed consent before completing the survey and were compensated US $2.70 for their time and effort. On average, the survey took 18.88 minutes to complete.

Ethics Approval

The study protocol was reviewed and approved by the University of Colorado Boulder Institutional Review Board (protocol 20-0441) and was conducted in accordance with the Declaration of Helsinki.

Study Measures

Respondents completed an investigator-developed quantitative survey that included a series of structured questionnaires assessing demographics, GSN app use, and sexual behaviors (refer to Multimedia Appendix 1 for complete survey measures).

Demographic Characteristics

Respondents were asked demographic questions, including their age, race, ethnicity, highest educational attainment, individual income, geographic location, and relationship status.

GSN App Use

Respondents were asked whether they used online dating apps, and if so, to indicate which apps they currently have profiles or accounts on (choosing all that applied). Response options included Grindr, Tinder, Scruff, Bumble, Hinge, GROWLr, Jack'd, Hornet, and other. Participants who selected *Other* were asked to list any additional apps on which they had profiles or accounts. The average weekly activity was assessed with the following item: "Please estimate the number of hours per week you spend on online dating apps such as the ones listed above."

Motivation for GSN App Use

Motivation for GSN app use was assessed using a measure developed by Goedel and Duncan [29]. Respondents who endorsed using GSN apps were asked, "Which best describes your reason for using these apps?" Response options included, "I want to 'kill time' when bored," "I want to make friends with other gay and bisexual men," "I want to meet other gay and bisexual men to date," "I want to find a boyfriend or other romantic partner," and "I want to meet other gay and bisexual men to have sex with."

Number of Sexual Partners

Respondents were asked, 'In the past year, with how many partners have you had anal sex?' After examining the response distribution, and consistent with prior studies [24,30], this variable was dichotomized before analysis to create a binary partner number variable (0=less than 3 sexual partners, 1=3 or more sexual partners).

Condomless Anal Intercourse

Past-year CAI was assessed with a single item asking respondents, "In the past year, with how many partners have you had unprotected anal sex?" After examining the response distribution, and again consistent with prior studies [22,24,30], this variable was dichotomized before analysis to create a binary CAI variable (0=have not had CAI in the past year, 1=have had CAI in the past year).

Sexualized Drug Use

Respondents were asked to indicate how frequently they consumed alcohol and cannabis before having sex by responding to the following two items: "In the past year, how much of the time did you drink alcohol before you had sexual intercourse?" and "In the past year, how much of the time did you smoke marijuana before you had sexual intercourse?" The response options for each item ranged from 1 (never) to 5 (always). In addition, respondents were given a list of 6 drugs commonly associated with *chemsex* (eg, ecstasy, poppers, and gamma-hydroxybutyrate) [31] and were asked to report whether they had ever used each drug before engaging in sexual intercourse. Items were summed to create a single score for recreational drug use before sex, ranging from 0 to 6.

Previous STI Diagnosis

Respondents were asked if they had ever been diagnosed with an STI other than HIV (0=have not been diagnosed, 1=have been diagnosed), and if so, which STI. Response options included chlamydia, genital herpes, gonorrhea, hepatitis, human papillomavirus, syphilis, trichomoniasis, and other. Participants who selected *Other* were asked to list any additional STIs they had previously been diagnosed with.

HIV Testing and Serostatus

Respondents were asked whether they had ever been tested for HIV (0=no, 1=yes), and if so, what their HIV status was (0=HIV negative, 1=HIV positive).

PrEP Use

PrEP use was assessed using a single item asking respondents, "PrEP is short for pre-exposure prophylaxis. It is a medication that HIV-negative individuals can take to prevent HIV. Do you use PrEP?" (0=do not use PrEP, 1=do use PrEP).

Sexual Sensation Seeking

Sexual sensation seeking was assessed using an abbreviated version of the Sexual Sensation Seeking scale [32], which assesses an individual's tendency to seek novel or risky sexual simulation. Respondents were asked to respond to 6 items (eg, "I like new and exciting sexual experiences and sensation") on a scale ranging from 1 (not at all like me) to 4 (very much like me). Items were averaged to create a single score ranging from 1 to 4, with higher scores indicating greater sexual sensation seeking (α=.78).

Statistical Analysis

All analyses were conducted using R (version 4.0.3). To compare GSN app users with nonusers on continuous outcomes of interest (eg, sexualized drug use), we ran a series of linear regressions with user status as the independent variable (nonuser=−0.5, GSN user=0.5). To compare GSN users with nonusers on categorical outcomes of interest (eg, past-year CAI), we ran a series of logistic regressions with user status as the independent variable. All regression analyses comparing GSN users with nonusers included relationship status (in a relationship=−0.5, not in a relationship=0.5) as a covariate, as relationship status differed significantly between the 2 groups (P<.001). Among GSN users, we conducted bivariate correlation tests to examine associations between hours spent per week on GSN apps and continuous outcomes of interest and point-biserial correlation tests to examine associations between hours spent per week on GSN apps and categorical outcomes of interest.

Finally, to examine the differential associations between behavior and GBM-specific versus sexuality nonspecific GSN apps, we conducted a series of exploratory analyses to investigate whether sexual risk behavior and substance use differed between those using *only* GBM-specific GSN apps (ie, Grindr, Scruff, GROWLr, Jack'd, and Hornet) and those using *only* sexuality nonspecific GSN apps (ie, Tinder, Bumble, and Hinge). Respondents who reported using both GBM-specific and sexuality nonspecific GSN apps (n=39) were excluded from the analyses. To compare exclusive users of GBM-specific apps with exclusive users of sexuality nonspecific apps on categorical outcomes of interest, we ran a series of logistic regressions with app preference (sexuality nonspecific only=−0.5, GBM-specific only=0.5) as the independent variable. To compare these users on the continuous outcomes of interest, we ran a series of 2-tailed independent samples *t* tests.

Results

Sample Characteristics

In total, 230 individuals completed the web-based survey. A total of 7 respondents self-reported their sexual orientation as heterosexual and were thus excluded. Data were examined for invalid survey response patterns (eg, failed attention checks, invariability in responses, and speeding), and no additional respondents were identified for exclusion. Thus, the final sample consisted of 223 respondents. The respondents were aged, on average, 31.90 (SD 10.06; range 21-78) years. The sample mostly consisted of non-Hispanic White participants (155/223, 69.5%) and included respondents from 40 states and the District of Columbia. GSN users were less likely to report being in a relationship (P<.001) and more likely to be a racial minority (P=.003) than nonusers. No other significant differences in demographics emerged between GSN users and nonusers (Table 1).

Nearly half (104/223, 47%) of the participants reported using GSN apps. Among GSN users, the most common reason for using these apps was to *kill time* when bored (43/104, 41.3%), followed by meeting other gay and bisexual men to have sex with (23/104, 22.1%), making friends with other gay and bisexual men (19/104, 18.3%), meeting other gay and bisexual men to date (10/104, 9.6%), and wanting to meet a boyfriend or other romantic partner (9/104, 8.7%). Respondents reported spending an average of 4.88 (SD 8.08) hours per week on these apps. The most common apps GSN users endorsed were Grindr (68/104, 65.4%), followed by Tinder (60/104, 57.7%), Scruff (27/104, 26.0%), Bumble (15/104, 14.4%), Hinge (14/104, 13.5%), GROWLr (11/104, 10.6%), Jack'd (8/104, 8.0%), and Hornet (4/104, 3.8%). Of 104 GSN users, 40 (38.5%) reported using 2 or more apps.

Table 1. Demographics overall and by user group.

Variable	Overall (N=223)	App users (n=104)	Nonusers (n=119)	*P* value
Age (years), mean (SD)	31.90 (10.06)	31.12 (8.53)	32.59 (11.22)	.28
Sexual orientation, n (%)				.99
Gay	100 (44.8)	47 (45.2)	53 (44.5)	
Bisexual	104 (46.6)	48 (46.2)	56 (47.1)	
Other	19 (8.5)	9 (8.7)	10 (8.4)	
Relationship status, n (%)				*<.001* [a]
Single	122 (54.7)	70 (67.3)	52 (43.7)	
Partnered (monogamous)	91 (40.8)	27 (26.0)	64 (53.8)	
Partnered (nonmonogamous)	10 (4.5)	7 (6.7)	3 (2.5)	
Race, n (%)				*.003*
White	159 (71.3)	65 (62.5)	94 (79.0)	
Black or African American	21 (9.4)	13 (12.5)	8 (6.7)	
Asian	17 (7.6)	14 (13.5)	3 (2.5)	
American Indian or Alaska Native	1 (0.4)	1 (1.0)	0 (0.0)	
Native Hawaiian or other Pacific Islander	0 (0.0)	0 (0.0)	0 (0.0)	
Two or more races	7 (3.1)	1 (1.0)	6 (5.0)	
Hispanic or Latinx—any race				.74
Hispanic or Latino	22 (9.9)	11 (10.6)	11 (9.2)	
Not Hispanic or Latino	201 (90.1)	93 (89.4)	108 (90.8)	
Education, n (%)				.63
Less than high school	2 (0.9)	0 (0.0)	2 (1.7)	
High school or general educational development	22 (9.9)	9 (8.7)	13 (10.9)	
Some college	63 (28.3)	29 (2.8)	34 (28.6)	
Associate degree or technical certification	18 (8.1)	10 (9.6)	8 (67.2)	
Bachelor’s degree	84 (37.7)	41 (39.4)	43 (36.1)	
Master’s degree	28 (12.6)	11 (10.6)	17 (14.3)	
Doctoral or professional degree	6 (2.7)	4 (3.8)	2 (1.7)	
Annual household income, n (%)				.35
< US $25,000	60 (26.9)	21 (20.2)	39 (32.8)	
US $25,000-$49,999	65 (29.1)	35 (33.7)	30 (25.2)	
US $50,000-$74,999	47 (21.1)	23 (22.1)	24 (20.2)	
US $75,000-$99,999	24 (10.8)	13 (12.5)	11 (9.2)	
US $100,000-$149,999	19 (8.5)	9 (8.7)	10 (8.4)	
> US $150,000	8 (3.6)	3 (2.9)	5 (4.2)	
Location of residence, n (%)				.46
Rural	24 (11.8)	10 (9.6)	14 (11.8)	
Suburban	110 (49.3)	48 (46.2)	62 (52.1)	
Urban	89 (39.9)	46 (44.2)	43 (36.1)	

[a]Significant *P* values are italicized.

GSN App Use, HIV and STI Prevalence, and Sexual Behavior

The zero-order correlations between sexual risk and protective behaviors are presented in Table 2.

Descriptive statistics for sexual behavior overall and by user group, as well as results of the regression analyses, are presented in Tables 3 and 4. Controlling for relationship status, GSN users reported higher levels of sexual sensation seeking than nonusers. GSN users also reported more frequent use of recreational drugs before sex, alcohol use before sex, and cannabis use before sex compared with nonusers. GSN users were more than 5 times more likely to report past-year CAI, 8 times more likely to report 3 or more sexual partners in the past year, and 3 times more likely to report a past STI diagnosis than nonusers. GSN users were also more likely than nonusers to report having ever taken an HIV test and using PrEP. HIV seropositivity rates were not significantly different between GSN users and nonusers.

Among GSN users (n=105), hours spent per week on GSN apps were positively associated with sexual sensation seeking (r=0.28; P=.003); frequency of recreational drug use before sex (r=0.25; P=.01); and the likelihood of past-year CAI (r=0.37; P<.001), 3 or more sexual partners in the previous year (r=0.38; P<.001), a past STI diagnosis (r=0.31; P=.001), HIV testing (r=0.26; P=.01), and PrEP use (r=0.44; P<.001). Hours spent on GSN apps were not associated with alcohol use before sex, cannabis use before sex, or HIV serostatus (Ps>.47).

Table 2. Zero-order correlations between sexual risk and protective behaviors.

Variable	1	2	3	4	5	6	7	8	9
Past-year condomless anal intercourse									
r	—[a]	—	—	—	—	—	—	—	—
P value	—	—	—	—	—	—	—	—	—
Three or more sexual partners in the past year									
r	0.35	—	—	—	—	—	—	—	—
P value	<.001	—	—	—	—	—	—	—	—
Previous sexually transmitted infection diagnosis									
r	0.31	0.25	—	—	—	—	—	—	—
P value	<.001	<.001	—	—	—	—	—	—	—
Ever been tested for HIV									
r	0.41	0.25	0.29	—	—	—	—	—	—
P value	<.001	<.001	<.001	—	—	—	—	—	—
HIV+ serostatus									
r	0.13	0.06	0.39	0.14	—	—	—	—	—
P value	.28	.76	<.001	<.05	—	—	—	—	—
Pre-exposure prophylaxis use									
r	0.26	0.48	0.34	0.22	0.01	—	—	—	—
P value	<.001	<.001	<.001	<.001	.79	—	—	—	—
Alcohol use before sex									
r	0.28	0.18	0.07	0.22	−0.08	0.12	—	—	—
P value	<.001	<.01	.31	<.001	.10	.08	—	—	—
Cannabis use before sex									
r	0.18	0.15	0.13	0.20	0.02	0.13	0.39	—	—
P value	<.01	<.05	.05	<.01	.94	.06	<.001	—	—
Recreational drug use before sex									
r	0.35	0.30	0.38	0.26	0.41	0.17	0.16	0.21	—
P value	<.001	<.001	<.001	<.001	<.001	<.05	<.05	<.01	—
Sexual sensation seeking									
r	0.55	0.29	0.37	0.33	0.21	0.19	0.33	0.28	0.44
P value	<.001	<.001	<.001	<.001	<.01	<.01	<.001	<.001	<.001

[a]Not applicable.

Table 3. Logistic regression analyses of geosocial networking app users versus nonusers.

Variable[a]	Overall (N=223)	App users (n=104)	Nonusers (n=119)	Adjusted odds ratio (95% CI)	*P* value
Past-year condomless anal intercourse, n (%)	103 (46.2)	64 (61.5)	39 (32.8)	5.46 (2.60-11.48)	*<.001* [b]
Three or more sexual partners in the past year, n (%)	53 (23.8)	44 (42.3)	9 (7.6)	8.26 (3.71-18.40)	*<.001*
Previous sexually transmitted infection diagnosis, n (%)	44 (19.7)	30 (28.8)	14 (11.8)	3.40 (1.62-7.11)	*.001*
Ever been tested for HIV, n (%)	145 (65.0)	75 (72.1)	70 (58.8)	3.00 (1.58-5.69)	*<.001*
HIV+serostatus, n (%)	8 (3.6)	6 (5.8)	2 (1.7)	1.73 (0.31-9.55)	.53
Pre-exposure prophylaxis use, n (%)	24 (10.8)	17 (16.3)	7 (5.9)	2.90 (1.12-7.56)	*.03*

[a]All regression models comparing GSN users and nonusers controlled for participant relationship status.

[b]Significant *P* values are italicized.

Table 4. Linear regression analyses of geosocial networking app users versus nonusers.

Variable[a]	Overall (N=223)	App users (n=104)	Nonusers (n=119)	Coefficient (*B*)	*t* test (*df*)	*P* value
Alcohol use before sex, mean (SD)	1.90 (1.02)	2.16 (1.06)	1.67 (0.93)	0.62	4.59 (220)	*<.001* [b]
Cannabis use before sex, mean (SD)	1.52 (0.87)	1.65 (0.97)	1.40 (0.75)	0.34	2.84 (220)	*.01*
Recreational drug use before sex, mean (SD)	0.65 (1.16)	0.90 (1.39)	0.43 (0.87)	0.53	3.30 (220)	*.001*
Sexual sensation seeking, mean (SD)	2.19 (0.70)	2.41 (0.68)	1.99 (0.65)	0.52	5.75 (220)	*<.001*

[a]All regression models comparing GSN users and nonusers controlled for participant relationship status.

[b]Significant *P* values are italicized.

GBM-Specific Versus Sexuality Nonspecific GSN Apps

Among GSN users, 38 respondents reported using *only* GBM-specific GSN apps (ie, Grindr, Scruff, GROWLr, Jack'd, and Hornet), whereas 27 respondents reported using *only* sexuality nonspecific GSN apps (ie, Tinder, Bumble, and Hinge). GSN users who only endorsed using only GBM-specific GSN apps spent more hours per week on GSN apps than GSN users who only endorsed using sexuality nonspecific GSN apps (t_{63}=2.34; *P*=.02). The most common reason for using GBM-specific apps was to meet other gay and bisexual men to have sex with (17/38, 45%), whereas the most common reason for using sexuality nonspecific apps was to kill time when bored (14/27, 52%).

Exclusive users of GBM-specific apps reported higher levels of sexual sensation seeking than exclusive users of sexuality nonspecific apps (t_{128}=5.14; *P*<.001). Exclusive users of GBM-specific apps also reported more frequent recreational drug use before sex (t_{63}=2.54; *P*=.01) and were more likely to report past-year CAI (OR 16.10, 95% CI 4.46-58.14; *P*<.001), 3 or more sexual partners in the previous year (OR 4.89, 95% CI 1.53-15.61; *P*=.004), a previous STI diagnosis (OR 26.00, 95% CI 3.20-211.49; *P*=.002), and HIV testing (OR 6.13, 95% CI 1.83-20.47; *P*=.003). Alcohol use before sex, cannabis use before sex, PrEP use, and HIV rates were similar between both the groups (*P*>.11).

Discussion

Principal Findings

As GBM continue to bear a disproportionate burden of the HIV and STI epidemic [1], it is important to examine how contextual factors, such as the proliferation of mobile dating apps, may shape STI risk within this community. This study suggests that GSN app use—dichotomous user versus nonuser status and time spent on apps among users—is associated with higher rates of sexual risk in GBM across a range of outcome measures. Conversely, the findings also suggest that GSN app use is associated with increased odds of engaging in health *protective* behaviors. Furthermore, among GSN users, we found that exclusive users of GBM-specific apps (eg, Grindr), as opposed to sexuality nonspecific apps (eg, Tinder), reported greater sexual risk taking despite similar rates of PrEP use.

Recent reviews have called for more studies to compare health risk and protective behaviors between GSN users and nonusers, as much of the extant literature consists of GBM samples recruited directly from GSN apps, which inherently limits findings to the prevalence of sexual risk among GSN users [6,14]. In our study that directly compared GSN using GBM with GSN nonusing GBM, we found that although GSN app use was associated with increased odds of a previous STI diagnosis, past-year CAI, and 3 or more sexual partners in the preceding year, it was also related to greater engagement in STI risk reduction strategies (ie, PrEP use and HIV testing). This is

encouraging, as it suggests that users may recognize the inherent risks associated with sexual behavior and actively engage with strategies for risk mitigation and prevention. Furthermore, our finding that GSN app use was associated with greater levels of sexualized drug use contributes to the emerging literature on this topic [15]. As there are well-documented event-level associations between substance use, sexual risk taking, and subsequent STI and HIV infection [33-35], this association—and the mechanisms behind it—warrants further exploration. For instance, GSN app use may reflect greater integration with GBM communities where sexualized drug use is normative [12,13]. Interventions and policy strategies aimed at reducing this practice among GBM should integrate both HIV and substance use education [12,13], and GSN apps could serve as a unique platform for disseminating these harm reduction initiatives.

In addition to comparing GSN users with nonusers, we provide a more nuanced exploration of the risks associated with GBM-specific (vs sexuality nonspecific) app use, which has been overlooked in the previous literature. Our findings indicate that men who used only GBM-specific apps (vs sexuality nonspecific apps) were more likely to report past-year CAI, 3 or more sexual partners in the preceding year, a previous STI diagnosis, and more frequent recreational drug use before sex, suggesting that the use of GBM-specific apps is associated with a higher risk user profile than sexuality nonspecific apps. The higher rates of risk behavior associated with GBM-specific apps may be driven by differences in motives for app use (eg, sexual partner seeking among GBM-specific app users vs killing time for users of sexuality nonspecific apps). This aligns with prior evidence suggesting that GBM-specific apps (eg, Grindr and Jack'd) are used primarily for *hookups* rather than dating [14]. Interestingly, despite greater levels of risk among GBM using only GBM-specific apps, the rates of PrEP use did not significantly differ between the groups; however, PrEP uptake was still low across both groups (13%). It is important to note that advertisements for PrEP are more common on GBM-specific dating sites, suggesting that further evaluation of the efficacy of such advertisements would be informative. Interestingly, although the rates of PrEP use were similar among app users, men who used only GBM-specific apps were more likely to report HIV testing than men who used only sexuality nonspecific apps. This may be due, in part, to the fact that GBM-specific apps encourage users to report their HIV serostatus as well as the date of their last HIV test.

Limitations

Despite the novel contributions made to the existing literature, this study has several limitations. Among GSN users, we were unable to determine whether self-reported risk behaviors occurred with partners met through GSN apps. Future studies should examine risk behaviors with partners who met online compared with those who met offline. Our self-report measures of sexual risk and protective behaviors also lacked nuance. For instance, although CAI is a risk factor for STIs, our measure may not accurately reflect respondents' personal levels of HIV risk, as it did not assess whether respondents who engaged in CAI were having sex with a seroconcordant partner, were regularly taking PrEP at the time of intercourse, or had an undetectable viral load (if HIV positive). Furthermore, our measure of PrEP use was a single yes-no question; however, evidence suggests that daily or almost-daily (4 or more pills per week) adherence to PrEP is necessary to maintain its efficacy [36]. In addition, this study's data regarding protective behaviors were limited to dichotomous measures of lifetime prevalence of STI and HIV testing and diagnoses, which did not align with the time frames for measures of app use and sexual risk behavior. Future studies should use more detailed self-report measures of risk and protective behaviors, such as the Timeline Follow-Back [37,38]. In addition, although our sample was diverse in terms of age (21-78 years) and geographic location (40 US states and the District of Columbia; urban, suburban, and rural areas), participants were predominantly White, whereas GBM of color—particularly Latino and Black GBM—are at far higher risk of HIV and other STI than White GBM [39]. Participants were also recruited as a convenience sample, which may limit the generalizability of our findings. Finally, this study was cross-sectional in nature. As such, we are unable to determine whether the relationships reported are causal, and if so, the directionality of these effects. For instance, although it is possible that GSN app use may *causally* influence the sexual risk behavior of GBM, it is also possible that GBM with a greater propensity for risk taking may self-select into these web-based environments. The fact that GSN app users reported higher levels of sexual sensation seeking lends credence to the self-selection hypothesis. Future prospective studies should examine how risk levels change in response to the initiation of GSN app use. Regardless of whether GSN app use causally influences risk taking, the fact remains that risk behaviors are elevated among GBM using these apps. Thus, these apps represent an important venue for targeted public health communications aimed at reducing STI risk.

Conclusions

To reduce disparities in STI and HIV infection and transmission rates, it is critical for researchers, clinicians, and public health officials to maintain an up-to-date awareness of contextual determinants of STI risk that are specific to high-risk populations. The findings of our study suggest that GSN apps, which have played an important role in reducing barriers and facilitating partner seeking among sexual minorities in the 10 years since their inception [9,10], may also be a useful pathway for evaluating and reducing STI risk in GBM. At the individual level, clinicians should ask about the use of dating apps when discussing patient history, given the relationship between the use of GSN apps and risky sexual behaviors. On a larger scale, GSN apps may be a useful tool for public health messaging and STI risk reduction interventions that address individual- and societal-level factors driving sexual risk among this population. The use of Grindr as a platform for PrEP advertisements in recent years is a promising start, but outreach efforts should be increased to reach GBM who prefer to use other GBM-specific apps (eg, Jack'd) or sexuality nonspecific apps (eg, Tinder), and further research is needed to evaluate the efficacy of such efforts.

Acknowledgments

This work was supported by a University of Colorado Boulder Beverly Sears Graduate Student Research Grant (principal investigator: LPG). LPG was funded by the National Science Foundation Graduate Research Fellowship (DGE 1650115).

Authors' Contributions

LPG was responsible for the study design, overseeing all aspects of study execution, and conceptualizing and performing data analysis. LPG and EBK led the interpretation and write up of the findings. ADB contributed to the conception and design, analysis and interpretation, drafting, and final approval. All authors critically reviewed the content and approved the final version for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Web-based survey measures.
[DOCX File , 26 KB - formative_v6i6e35548_app1.docx]

References

1. Diagnoses of HIV infection in the United States and dependent areas, 2018. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillance-report-2018-updated-vol-31.pdf [accessed 2021-12-08]
2. Sexual transmission and viral hepatitis. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/hepatitis/populations/stds.htm [accessed 2021-12-08]
3. Centers for Disease Control and Prevention. 2021. URL: https://www.cdc.gov/std/syphilis/stdfact-msm-syphilis.htm [accessed 2021-12-08]
4. Sexually Transmitted Disease Surveillance 2019. Centers for Disease Control and Prevention. 2021. URL: https://www.cdc.gov/std/statistics/2019/default.htm [accessed 2021-12-08]
5. Substance Abuse and Mental Health Services Administration. 2012. URL: https://store.samhsa.gov/sites/default/files/d7/priv/sma12-4684.pdf [accessed 2021-12-08]
6. Zou H, Fan S. Characteristics of men who have sex with men who use smartphone geosocial networking applications and implications for HIV interventions: a systematic review and meta-analysis. Arch Sex Behav 2017 May;46(4):885-894. [doi: 10.1007/s10508-016-0709-3] [Medline: 27043837]
7. Curry D. Dating App Revenue and Usage Statistics (2021). Business of Apps. 2022 Mar 4. URL: https://www.businessofapps.com/data/dating-app-market/ [accessed 2021-12-08]
8. Meisenzahl M. These charts from Match Group show more people are turning to online dating during the pandemic. Business Insider. 2020 Aug 6. URL: https://www.businessinsider.com/tinder-hinge-match-group-dating-apps-more-users-coronavirus-2020-8 [accessed 2021-12-08]
9. Brown A. Lesbian, gay and bisexual online daters report positive experiences – but also harassment. Pew Research Center. 2020 Apr 9. URL: https://www.pewresearch.org/fact-tank/2020/04/09/lesbian-gay-and-bisexual-online-daters-report-positive-experiences-but-also-harassment/ [accessed 2021-12-08]
10. Jaspal R. Gay men's construction and management of identity on Grindr. Sex Cult 2017;21(1):187-204. [doi: 10.1007/s12119-016-9389-3]
11. Liau A, Millett G, Marks G. Meta-analytic examination of online sex-seeking and sexual risk behavior among men who have sex with men. Sex Transm Dis 2006 Sep;33(9):576-584. [doi: 10.1097/01.olq.0000204710.35332.c5] [Medline: 16540884]
12. Chan AS, Tang PM. Application of novel psychoactive substances: Chemsex and HIV/AIDS policies among men who have sex with men in Hong Kong. Front Psychiatry 2021;12:680252 [FREE Full text] [doi: 10.3389/fpsyt.2021.680252] [Medline: 34335329]
13. Kong TS, Laidler KJ. The paradox for chem-fun and gay men: a neoliberal analysis of drugs and HIV/AIDS policies in Hong Kong. J Psychoactive Drugs 2020;52(1):77-85. [doi: 10.1080/02791072.2019.1568648] [Medline: 30704363]
14. Wang H, Zhang L, Zhou Y, Wang K, Zhang X, Wu J, et al. The use of geosocial networking smartphone applications and the risk of sexually transmitted infections among men who have sex with men: a systematic review and meta-analysis. BMC Public Health 2018;18(1):1178 [FREE Full text] [doi: 10.1186/s12889-018-6092-3] [Medline: 30326887]
15. Goedel WC, Duncan DT. Contextual factors in geosocial-networking smartphone application use and engagement in condomless anal intercourse among gay, bisexual, and other men who have sex with men who use Grindr. Sex Health 2016 Nov;13(6):549-554. [doi: 10.1071/SH16008] [Medline: 27712610]

16. Obarska K, Szymczak K, Lewczuk K, Gola M. Threats to mental health facilitated by dating applications use among men having sex with men. Front Psychiatry 2020 Nov 13;11:584548 [FREE Full text] [doi: 10.3389/fpsyt.2020.584548] [Medline: 33281645]
17. Phillips G, Grov C, Mustanski B. Engagement in group sex among geosocial networking mobile application-using men who have sex with men. Sex Health 2015 Nov;12(6):495-500 [FREE Full text] [doi: 10.1071/SH15108] [Medline: 26255156]
18. Landovitz RJ, Tseng CH, Weissman M, Haymer M, Mendenhall B, Rogers K, et al. Epidemiology, sexual risk behavior, and HIV prevention practices of men who have sex with men using GRINDR in Los Angeles, California. J Urban Health 2013 Aug;90(4):729-739 [FREE Full text] [doi: 10.1007/s11524-012-9766-7] [Medline: 22983721]
19. Rice E, Holloway I, Winetrobe H, Rhoades H, Barman-Adhikari A, Gibbs J, et al. Sex risk among young men who have sex with men who use Grindr, a smartphone geosocial networking application. J AIDS Clinic Res 2012 Aug;01(S4):729-739. [doi: 10.4172/2155-6113.S4-005]
20. Lehmiller JJ, Ioerger M. Social networking smartphone applications and sexual health outcomes among men who have sex with men. PLoS One 2014 Jan 23;9(1):e86603 [FREE Full text] [doi: 10.1371/journal.pone.0086603] [Medline: 24466166]
21. Ogunbajo A, Lodge 2nd W, Restar AJ, Oginni OA, Iwuagwu S, Williams R, et al. Correlates of geosocial networking applications (GSN Apps) usage among gay, bisexual, and other men who have sex with men in Nigeria, Africa. Arch Sex Behav 2021 Oct;50(7):2981-2993 [FREE Full text] [doi: 10.1007/s10508-020-01889-3] [Medline: 33844118]
22. Tang W, Best J, Zhang Y, Liu FY, Tso LS, Huang S, et al. Gay mobile apps and the evolving virtual risk environment: a cross-sectional online survey among men who have sex with men in China. Sex Transm Infect 2016 Nov;92(7):508-514 [FREE Full text] [doi: 10.1136/sextrans-2015-052469] [Medline: 27288414]
23. Wang H, Zhang J, Chu Z, Hu Q, Dong W, Huang X, et al. Risk-taking behaviors and adherence to HIV pre-exposure prophylaxis in users of geosocial networking apps: real-world, multicenter study. J Med Internet Res 2020 Oct 14;22(10):e22388 [FREE Full text] [doi: 10.2196/22388] [Medline: 33052132]
24. Xu J, Yu H, Tang W, Leuba SI, Zhang J, Mao X, et al. The effect of using geosocial networking apps on the HIV incidence rate among men who have sex with men: eighteen-month prospective cohort study in Shenyang, China. J Med Internet Res 2018 Dec 21;20(12):e11303 [FREE Full text] [doi: 10.2196/11303] [Medline: 30578225]
25. Bien CH, Best JM, Muessig KE, Wei C, Han L, Tucker JD. Gay apps for seeking sex partners in China: implications for MSM sexual health. AIDS Behav 2015 Jun;19(6):941-946 [FREE Full text] [doi: 10.1007/s10461-014-0994-6] [Medline: 25572834]
26. Hoenigl M, Little SJ, Grelotti D, Skaathun B, Wagner GA, Weibel N, et al. Grindr users take more risks, but are more open to Human Immunodeficiency Virus (HIV) pre-exposure prophylaxis: could this dating app provide a platform for HIV prevention outreach? Clin Infect Dis 2020 Oct 23;71(7):e135-e140 [FREE Full text] [doi: 10.1093/cid/ciz1093] [Medline: 31677383]
27. Macapagal K, Kraus A, Moskowitz DA, Birnholtz J. Geosocial networking application use, characteristics of app-met sexual partners, and sexual behavior among sexual and gender minority adolescents assigned male at birth. J Sex Res 2020 Oct;57(8):1078-1087 [FREE Full text] [doi: 10.1080/00224499.2019.1698004] [Medline: 31846592]
28. Palan S, Schitter C. Prolific.ac—a subject pool for online experiments. J Behav Exp Finance 2018 Mar;17:22-27. [doi: 10.1016/j.jbef.2017.12.004]
29. Goedel WC, Duncan DT. Geosocial-networking app usage patterns of gay, bisexual, and other men who have sex with men: survey among users of Grindr, a mobile dating app. JMIR Public Health Surveill 2015 May 8;1(1):e4 [FREE Full text] [doi: 10.2196/publichealth.4353] [Medline: 27227127]
30. Balaji AB, Bowles KE, Hess KL, Smith JC, Paz-Bailey G, NHBS study group. Association between enacted stigma and HIV-related risk behavior among MSM, National HIV Behavioral Surveillance System, 2011. AIDS Behav 2017 Jan;21(1):227-237. [doi: 10.1007/s10461-016-1599-z] [Medline: 27830344]
31. Bourne A, Reid D, Hickson F, Torres-Rueda S, Weatherburn P. Illicit drug use in sexual settings ('chemsex') and HIV/STI transmission risk behaviour among gay men in South London: findings from a qualitative study. Sex Transm Infect 2015 Dec;91(8):564-568. [doi: 10.1136/sextrans-2015-052052] [Medline: 26163510]
32. Kalichman SC, Rompa D. Sexual sensation seeking and Sexual Compulsivity Scales: reliability, validity, and predicting HIV risk behavior. J Pers Assess 1995 Dec;65(3):586-601. [doi: 10.1207/s15327752jpa6503_16] [Medline: 8609589]
33. Heath J, Lanoye A, Maisto SA. The role of alcohol and substance use in risky sexual behavior among older men who have sex with men: a review and critique of the current literature. AIDS Behav 2012 Apr;16(3):578-589 [FREE Full text] [doi: 10.1007/s10461-011-9921-2] [Medline: 21390534]
34. Mustanski B. Moderating effects of age on the alcohol and sexual risk taking association: an online daily diary study of men who have sex with men. AIDS Behav 2008 Jan;12(1):118-126. [doi: 10.1007/s10461-007-9335-3] [Medline: 18034298]
35. Ritchwood TD, DeCoster J, Metzger IW, Bolland JM, Danielson CK. Does it really matter which drug you choose? An examination of the influence of type of drug on type of risky sexual behavior. Addict Behav 2016 Sep;60:97-102 [FREE Full text] [doi: 10.1016/j.addbeh.2016.03.022] [Medline: 27104799]
36. Spinner CD, Boesecke C, Zink A, Jessen H, Stellbrink HJ, Rockstroh JK, et al. HIV pre-exposure prophylaxis (PrEP): a review of current knowledge of oral systemic HIV PrEP in humans. Infection 2016 Apr;44(2):151-158. [doi: 10.1007/s15010-015-0850-2] [Medline: 26471511]

37. Pachankis JE, McConocha EM, Reynolds JS, Winston R, Adeyinka O, Harkness A, et al. Project ESTEEM protocol: a randomized controlled trial of an LGBTQ-affirmative treatment for young adult sexual minority men's mental and sexual health. BMC Public Health 2019 Aug 09;19(1):1086 [FREE Full text] [doi: 10.1186/s12889-019-7346-4] [Medline: 31399071]
38. Weinhardt LS, Carey MP, Maisto SA, Carey KB, Cohen MM, Wickramasinghe SM. Reliability of the timeline follow-back sexual behavior interview. Ann Behav Med 1998;20(1):25-30 [FREE Full text] [doi: 10.1007/BF02893805] [Medline: 9755348]
39. HIV and gay and bisexual men. Centers for Disease Control and Prevention. 2021. URL: https://www.cdc.gov/hiv/group/msm/index.html [accessed 2021-12-08]

Abbreviations

CAI: condomless anal intercourse
GBM: gay, bisexual, and other men who have sex with men
GSN: geosocial networking
LGB: lesbian, gay, and bisexual
MSM: men who have sex with men
PrEP: pre-exposure prophylaxis
STI: sexually transmitted infection

Edited by A Mavragani; submitted 08.12.21; peer-reviewed by K Card, ASW Chan ; comments to author 07.03.22; revised version received 14.03.22; accepted 16.03.22; published 13.06.22.

Please cite as:
Gibson LP, Kramer EB, Bryan AD
Geosocial Networking App Use Associated With Sexual Risk Behavior and Pre-exposure Prophylaxis Use Among Gay, Bisexual, and Other Men Who Have Sex With Men: Cross-sectional Web-Based Survey
JMIR Form Res 2022;6(6):e35548
URL: https://formative.jmir.org/2022/6/e35548
doi:10.2196/35548
PMID:35699992

Original Paper

Acceptance, Barriers, and Facilitators to Implementing Artificial Intelligence–Based Decision Support Systems in Emergency Departments: Quantitative and Qualitative Evaluation

Ryo Fujimori[1,2], MD; Keibun Liu[2,3], MD, PhD; Shoko Soeno[2,4], MD; Hiromu Naraba[2,5], MD; Kentaro Ogura[1,2]; Konan Hara[2,6], MD, PhD; Tomohiro Sonoo[2,5], MD; Takayuki Ogura[7], MD, PhD; Kensuke Nakamura[5], MD, PhD; Tadahiro Goto[2], MPH, MD, PhD

[1]Faculty of Medicine, The University of Tokyo, Tokyo, Japan

[2]TXP Medical Co Ltd, Tokyo, Japan

[3]Critical Care Research Group, The Prince Charles Hospital, Brisbane, Australia

[4]Department of Palliative Care, Southern Tohoku General Hospital, Fukushima, Japan

[5]Department of Emergency and Critical Care Medicine, Hitachi General Hospital, Ibaraki, Japan

[6]Department of Economics, University of Arizona, Tucson, AZ, United States

[7]Department of Emergency and Critical Care Medicine, Saiseikai Utsunomiya Hospital, Tochigi, Japan

Corresponding Author:
Ryo Fujimori, MD
Faculty of Medicine
The University of Tokyo
7-3-1, Hongo, Bunkyo-ku
Tokyo, 1138655
Japan
Phone: 81 08095261328
Email: fujimori0203@gmail.com

Abstract

Background: Despite the increasing availability of clinical decision support systems (CDSSs) and rising expectation for CDSSs based on artificial intelligence (AI), little is known about the acceptance of AI-based CDSS by physicians and its barriers and facilitators in emergency care settings.

Objective: We aimed to evaluate the acceptance, barriers, and facilitators to implementing AI-based CDSSs in the emergency care setting through the opinions of physicians on our newly developed, real-time AI-based CDSS, which alerts ED physicians by predicting aortic dissection based on numeric and text information from medical charts, by using the Unified Theory of Acceptance and Use of Technology (UTAUT; for quantitative evaluation) and the Consolidated Framework for Implementation Research (CFIR; for qualitative evaluation) frameworks.

Methods: This mixed methods study was performed from March to April 2021. Transitional year residents (n=6), emergency medicine residents (n=5), and emergency physicians (n=3) from two community, tertiary care hospitals in Japan were included. We first developed a real-time CDSS for predicting aortic dissection based on numeric and text information from medical charts (eg, chief complaints, medical history, vital signs) with natural language processing. This system was deployed on the internet, and the participants used the system with clinical vignettes of model cases. Participants were then involved in a mixed methods evaluation consisting of a UTAUT-based questionnaire with a 5-point Likert scale (quantitative) and a CFIR-based semistructured interview (qualitative). Cronbach α was calculated as a reliability estimate for UTAUT subconstructs. Interviews were sampled, transcribed, and analyzed using the MaxQDA software. The framework analysis approach was used during the study to determine the relevance of the CFIR constructs.

Results: All 14 participants completed the questionnaires and interviews. Quantitative analysis revealed generally positive responses for user acceptance with all scores above the neutral score of 3.0. In addition, the mixed methods analysis identified two significant barriers (System Performance, Compatibility) and two major facilitators (Evidence Strength, Design Quality) for implementation of AI-based CDSSs in emergency care settings.

Conclusions: Our mixed methods evaluation based on theoretically grounded frameworks revealed the acceptance, barriers, and facilitators of implementation of AI-based CDSS. Although the concern of system failure and overtrusting of the system could be barriers to implementation, the locality of the system and designing an intuitive user interface could likely facilitate the use of optimal AI-based CDSS. Alleviating and resolving these factors should be key to achieving good user acceptance of AI-based CDSS.

(JMIR Form Res 2022;6(6):e36501) doi:10.2196/36501

KEYWORDS

clinical decision support system; preimplementation; qualitative; mixed methods; artificial intelligence; emergency medicine; CDSS; computerized decision; computerized decision support system; AI; AI-based; CFIR; quantitative analysis

Introduction

Clinical decision support systems (CDSSs) are computerized tools that are developed to assist clinicians in their decision-making processes with the ultimate goal of improving patient outcomes [1]. CDSSs support clinicians by employing various functions, including diagnostic support, disease management, prescription control, and drug control [1]. These CDSSs have been continuously developed over the past years and have become increasingly available in all areas of health care, including the emergency care setting [2-4]. In addition, the rapid development of computer science has led to advancements in artificial intelligence (AI)-based CDSSs [5], and the development of electronic health record (EHR) systems has enabled researchers to advance models and systems in the health care setting [6]. In the emergency department, physicians need to maximize their performance in a limited amount of time to deal with the high urgency and severity of the patients' conditions. To address such needs of emergency physicians, multiple CDSSs, such as alert systems and diagnostic imaging support systems, have been developed [7,8].

Despite the increased availability of CDSSs, including AI-based CDSSs in emergency care settings, the use of these systems is limited and has yet to achieve widespread implementation [1,9]. Several studies, mostly those outside emergency care settings, have identified reasons for the low usage and/or effectiveness of CDSSs. For example, studies have attributed the lack of usability, lack of integration with host systems, lack of time to effectuate advice, and alert fatigue to the low usage of CDSSs [9-11]. In addition, as medical device approval is needed in Japan for most CDSSs, time and effort are needed for implementation. Despite such knowledge regarding the implementation of CDSSs, little is known about the acceptance, barriers, and facilitators of AI-based CDSSs in the emergency care setting.

To address this knowledge gap, we aimed to evaluate the acceptance, barriers, and facilitators to implementing AI-based CDSSs in the emergency care setting through the opinions of physicians on our newly developed, real-time AI-based CDSS, which alerts ED physicians by predicting aortic dissection based on numeric and text information from medical charts, by using the Unified Theory of Acceptance and Use of Technology (UTAUT; for quantitative evaluation) [12] and Consolidated Framework for Implementation Research (CFIR; for qualitative evaluation) frameworks [13].

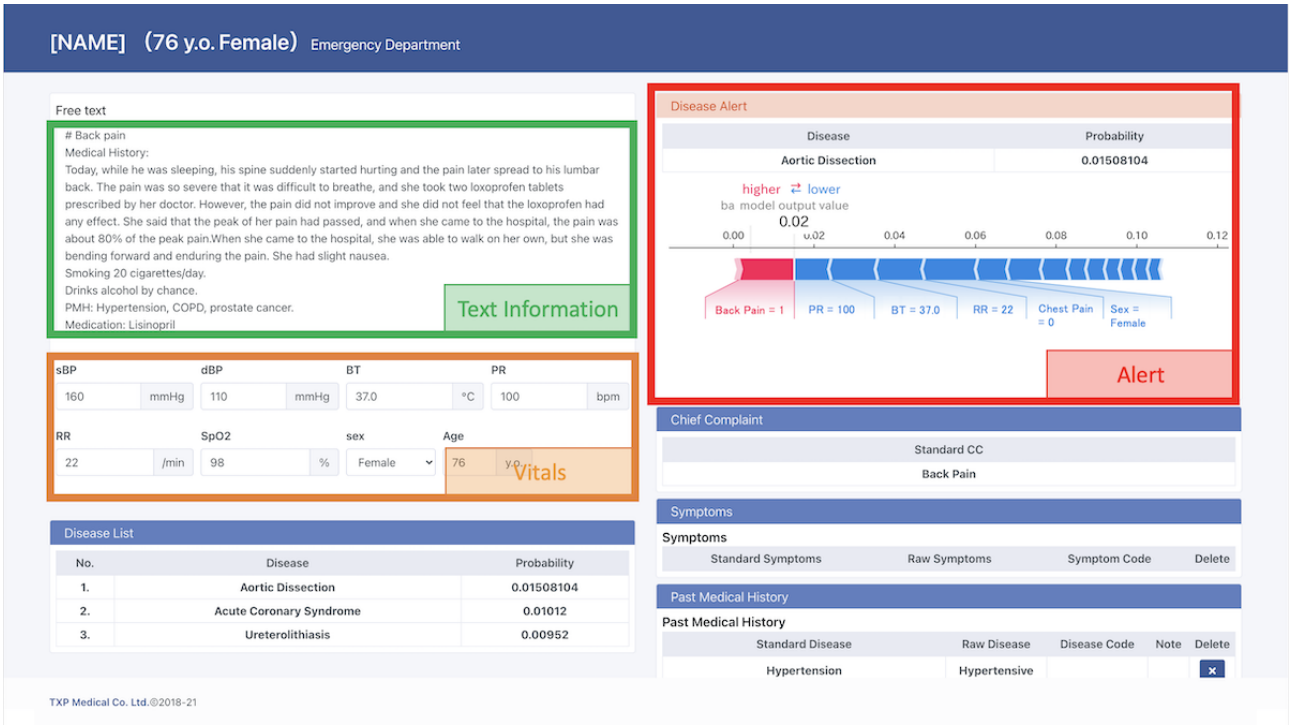
Methods

Study Design, Setting, and Participants

This is a mixed methods (ie, quantitative and qualitative), cross-sectional study of 14 physicians from two community, tertiary care hospitals in Japan (6 transitional year residents with 0-2 years of clinical experience, 5 emergency medicine residents with 3-5 years of clinical experience, and 3 emergency physicians with 5 years of clinical experience). The qualitative sampling used convenience sampling. This study was performed from March to April 2021. The physicians participated in a 1.5-hour session consisting of the following three sections: (1) a brief introduction of the system and the focus of the research, (2) actual use of the AI-based CDSS deployed on the internet with clinical vignettes of model cases, and (3) participation in a mixed methods evaluation consisting of a UTAUT-based questionnaire with a 5-point Likert scale and a CFIR-based semistructured interview. All interviews were moderated by two of the four research team members (RF, TG, KL, SS).

First, the moderator gave a brief introduction about the focus of the research and the CDSS, including how alerts are based on machine learning models. Verbal consent for participation in the research was obtained from each participant. In the second section, the participants used the AI-based CDSS deployed on the internet (Figure 1) with clinical vignettes of model cases created by two emergency physicians who authored this study (TG and KL). For this study, we implemented a machine learning–based CDSS for aortic dissection that consisted of an emergency alert system. We prepared both typical and atypical cases of the disease (Table S1 in Multimedia Appendix 1). Typical/atypical cases are defined by an author (RF) and confirmed by an emergency physician coauthor (TG). During the session, the moderator answered questions concerning the aim and background of the research, but questions on how to use the system and about the user interface were taken later in the interview to ensure that the explanations of the system were the same among participants. In the third section, the physicians participated in a mixed methods evaluation consisting of a UTAUT-based questionnaire with a 5-point Likert scale and a CFIR-based semistructured interview. All of these sessions were conducted using online video chat tools considering the infection risks of COVID-19.

Figure 1. A real-time clinical decision support system with Emergency Alert System for predicting aortic dissection based on numeric and text information from medical charts (eg, chief complaints, medical history, vital signs) organized using natural language processing.



The interviews were recorded using an audio recording device for ease of transcription and review. Data were transferred from the device following each interview and transcribed verbatim. The interview transcriptions were uploaded onto the MaxQDA 12 software (VERBI GmbH) for qualitative analysis [14]. All interviews were deidentified; a code was allocated to each interview, and personal identifiers were removed from the data. The codes were allocated based on the participant's medical experience: transitional year residents (TYR1 to TYR6), emergency medicine residents (EMR1 to EMR5), and emergency physicians (EP1 to EP3).

Proposed AI-Based CDSS

By using data from 27,550 emergency department patients from a tertiary care hospital in Japan, we first developed a real-time CDSS consisting of the Emergency Alert System, which notifies emergency department physicians by predicting aortic dissection based on numeric and text information from medical charts (eg, chief complaints, medical history, vital signs) organized using natural language processing (Figure 1). Aortic dissection was chosen as a pilot disease since it is an emergent, potentially fatal disease with numerous symptoms that can mimic non–life-threatening conditions [15]. The Emergency Alert System predicts the probability of the disease using real-time data input on the computer screen. The model used for prediction was developed using the XGBoost model [16] and hyperparameters were determined by 5-fold cross-validation to maximize the area under the receiver operator characteristic curve (AUROC). The AUROC of the model was 0.901 (95% CI 0.840-0.962). In an attempt to address anchoring bias while avoiding alert fatigue, the Emergency Alert System displays an alert when the probability of aortic dissection changes sharply (greater than 4 times the baseline risk) or exceeds the prespecified threshold of a predicted probability of 9.0%. In addition, to enhance the interpretability of the alert, the contribution of each feature (eg, chief complaint or systolic blood pressure) was calculated and was added to the alert on the screen as Shapley Additive exPlanations values [17]. The degree to which each feature contributed both positively and negatively was shown as a bar chart. To fully conduct the session online, this system was deployed on the internet via Amazon Web Services.

Theoretical Framework Selection for Quantitative and Qualitative Evaluation

We used the UTAUT model to quantitatively evaluate users' willingness to accept the proposed AI-based CDSS. The UTAUT model is composed of six main constructs that impact technology adoption: (1) effort expectancy, (2) performance expectancy, (3) social influence, (4) facilitating environment, (5) attitude toward using technology, and (6) behavioral intention [12]. We chose quantitative assessment for UTAUT because its components have relationships that have been determined by previous studies. Because this model does not consider the unique characteristics of the clinical setting (eg, limited time and human resources in the emergency department setting) [18], we further adopted qualitative research techniques based on the CFIR to further identify barriers and facilitators of the AI-based CDSS [13].

The CFIR was chosen because it is a relatively new framework that synthesizes prior research evidence into one consolidated framework with multiple constructs. In addition, the CFIR has flexibility in assessing implementation barriers and facilitators of research findings and innovations [13]. The CFIR consists of 39 constructs organized into 5 major domains found to influence the successful implementation of innovative programs. The domains assess the following characteristics of innovative programs: (1) intervention characteristics, (2) outer setting, (3)

inner setting, (4) characteristics of individuals, and (5) process [13].

Quantitative Analysis for UTAUT Questionnaire

The questionnaire for quantitative analysis was developed to investigate user attitudes toward the CDSS. A total of 23 questions were included in the questionnaire; of these, 21 represented concepts from the UTAUT and 2 questions were added for basic characteristics (age and sex). The phrasing of the questions was based on the original UTAUT article [19].

The questionnaire consisted of questions about the constructs of UTAUT: performance expectancy, effort expectancy, social influence, facilitating conditions, attitude toward using technology, and behavioral intention. Two to six questions were created to assess each domain, and each question was answered with a 5-point Likert scale. To confirm whether the UTAUT subconstructs are reliable in measuring the same construct, Cronbach α was calculated as a reliability estimate [20].

Qualitative Analysis for CFIR Questionnaires

To objectively report results, the qualitative analysis followed the Consolidated Criteria for Reporting Qualitative Research (COREQ) guidelines [21]. Interviews were conducted based on CFIR [22]. For this study, the CFIR domains aligned with the following entities: intervention characteristics (Emergency Alert System), outer setting (community, tertiary care hospitals), inner setting (emergency department), and characteristics of individuals (emergency clinicians who piloted the Emergency Alert System). The process domain, which describes how implementation should be enacted, was excluded because the system was in a preimplementation phase and it was thought to be irrelevant to the study. Overall, 16 constructs from 4 domains were selected for the semistructured interview (Table 1).

Table 1. Consolidated Framework for Implementation Research domains and constructs.

Domain	Constructs
I. Intervention characteristics[a]	
A	Intervention Source
B[a]	Evidence Strength and Quality
C[a]	Relative Advantage
D[a]	Adaptability
E	Trialability
F	Complexity
G[a]	Design Quality and Packaging
H	Cost
II. Outer setting[a]	
A[a]	Patient Needs and Resources
B	Cosmopolitanism
C	Peer Pressure
D[a]	External Policy and Incentives
III. Inner setting[a]	
A	Structural Characteristics
B	Networks and Communications
C[a]	Culture
D[a]	Implementation Climate
1[a]	Tension for Change
2[a]	Compatibility
3[a]	Relative Priority
4[a]	Organizational Incentives and Rewards
5	Goals and Feedback
6	Learning Climate
E[a]	Readiness for Implementation
1	Leadership Engagement
2[a]	Available Resources
3[a]	Access to Knowledge and Information
IV. Characteristics of individuals[a]	
A[a]	Knowledge and Beliefs about the Intervention
B	Self-efficacy
C	Individual Stage of Change
D	Individual Identification with Organization
E	Other Personal Attributes
V. Process	
A	Planning
B	Engaging

Domain	Constructs
C	Executing
1	Opinion Leaders
2	Formally Appointed Internal Implementation Leaders
3	Champions
4	External Change Agents
D	Reflecting and Evaluating

[a]The domains and constructs selected for the semistructured interview in this study.

The framework analysis approach [23,24] was used during the study to determine the relevance of the CFIR constructs. The framework analysis followed the 5-step process outlined by Richie and Spencer [25]: (1) familiarization, (2) identifying a thematic framework, (3) indexing, (4) charting, and (5) mapping/interpretation. The analysis was an ongoing iterative process. We conducted multiple reviews of the transcripts and video data to become familiar with the data (Step 1) and identify initial themes that were reflexive and interactive (Step 2). Analyses were initiated as soon as the first interview was completed and were continued concurrently with data collection to help determine when new information was no longer being generated from interviews. The team used the CFIR as the a priori framework, and the codes identified during the familiarization process were added to the CFIR. The codes also reflected relevant CFIR constructs across the 5 domains and were indexed to sections of the transcripts (Step 3). Sections of the transcripts were charted into themes and a summary matrix was organized with CFIR domains and constructs (Step 4). Two analysts reviewed the codes and associated themes multiple times to check for potential bias, to ensure they reflected participants' words and that their interpretation of the interviews was credible (Step 5).

Ethical Approval

We have confirmed with the ethics committee of TXP Medical Co Ltd that this study can be waived from ethical approval as it did not involve any patients.

Results

Overview

All 14 participants completed the questionnaires and interviews. The participant demographics and characteristics are shown in Table 2. All participants reported that they were unfamiliar or very unfamiliar with information technology. The UTAUT-based questionnaire lasted 3-7 minutes and the CFIR-based semistructured interview lasted 40-60 minutes. An interview of one participant was rescheduled due to network problems.

Table 2. Participant demographics (N=14).

Demographic	Participants, n (%)
Specialty	
Transitional year resident	6 (43)
Emergency medicine resident	5 (36)
Emergency physician	3 (21)
Gender	
Male	9 (64)
Female	5 (36)
Age group (years)	
20-29	7 (50)
30-39	6 (43)
40-49	1 (7)
Information technology familiarity	
Very familiar	0 (0)
Familiar	0 (0)
Neutral	0 (0)
Unfamiliar	3 (21)
Very unfamiliar	11 (79)

Quantitative Analysis

For the 6 UTAUT constructs investigated, Table 3 lists the mean and standard deviation of the Likert scale for each question and the Cronbach α reliability statistics for each concept. Although the Cronbach α values of Facilitating Conditions and Social Influence were less than .6, other constructs exhibited good reliability within the recommended range of Cronbach α>.60. The analysis revealed generally positive responses for user acceptance, with high scores on Attitude and Intention to Use (mean 3.43, SD 0.76). Although 79% (11/14) of the participants were not familiar with information technology, their perceived complexity of the system was low (mean 4.14, SD 0.72).

Table 3. Construct reliability and mean (SD) scores for the Unified Theory of Acceptance and Use of Technology–based questionnaires (5-point Likert scale).

Construct	Mean (SD)	Cronbach α
Performance Expectancy (PE)		.638
PE1: I would find the system useful in my job.	4.07 (0.73)	
PE2: Using the system enables me to accomplish tasks more quickly.	3.14 (0.66)	
PE3: Using the system improves the quality of the work I do.	3.57 (0.76)	
PE4: Using the system enhances my effectiveness on the job.	3.86 (0.66)	
PE5: If I use the system...My coworkers will perceive me as competent.	2.86 (0.95)	
Effort Expectancy (EE)		.690
EE1: My interaction with the system would be clear and understandable.	4.36 (0.63)	
EE2: I would find the system to be flexible to interact with.	3.57 (0.65)	
EE3: It would be easy for me to become skillful at using the system.	4.50 (0.52)	
EE4: Working with the system is so complicated, it is difficult to understand what is going on.	3.36 (0.50)	
EE5: Using the system involves too much time doing mechanical operations (eg, data input).	3.00 (0.96)	
EE6: My interaction with the system is clear and understandable.	3.86 (1.10)	
Social Influence (SI)		.499
SI1: People who influence my behavior think that I should use the system.	3.00 (0.96)	
SI2: Having the system is a status in my organization.	3.64 (1.45)	
Facilitating Conditions (FC)		.564
FC1: I have the knowledge necessary to use the system.	3.14 (1.03)	
FC2: Given the resources it takes to use the system, it would be easy for me to use the system.	4.21 (1.05)	
FC3: A specific person (or group) is available for assistance with system difficulties.	4.07 (0.83)	
Attitude Toward Using Technology (AT)		.760
AT1: Using the system is a bad/good idea.	4.14 (0.77)	
AT2: I have fun using the system.	3.36 (0.84)	
Behavior Intention (BI)		.740
BI1: I prefer to work with the system.	3.36 (0.84)	
BI2: I intend to use the system in the next 3 months.	3.43 (0.76)	

Qualitative Analysis

Overview

Table 4 presents the primary codes and count data differentiated by the CFIR domain and construct, as well as whether codes were barriers to or facilitators of Emergency Alert System implementation and adoption. Only relevant CFIR domains and constructs were coded and presented here. There were four key factors in the implementation of the system: Evidence Strength and Quality, Relative Advantage, Design Quality and Packaging, and Compatibility. Other factors influencing the implementation can be found in Table S2 in Multimedia Appendix 1.

Table 4. Primary codes and count data differentiated by Consolidated Framework for Implementation Research domain and construct.

Constructs	Barriers	Count	Facilitators	Count
I. Intervention Characteristics				
Evidence Strength and Quality				
	Distrust of the results	1	Sample size was enough for developing the model	9
			Local trends of disease	1
Relative Advantage				
	Unnecessary for experienced emergency physicians	1	Potential to reduce misdiagnoses	1
	Unnecessary for typical cases	6	More useful than diagnostic rules	2
	Alternatives to the system are enough	2	Never seen similar systems	8
	Can bias physicians' decision-making	2	Can aid diagnosis for difficult cases	6
	Limited use cases	1	Good for information sharing	1
			Useful for unexperienced physicians	3
III. Inner Settings				
Design Quality and Packaging				
	Unable to find when the system shows alerts	1	Easy to use and not interruptive	10
	Potentially distracting for comorbidities	1	Summary board was informative	1
			Real-time alerts were intuitive	2
Compatibility				
	Anxious if the system is not working properly	1	Easily integrated with existing workflow	14
	Affects typing speed	1		
	Fear of system failure	1		

Evidence Strength and Quality

Nine key informants emphasized that the sample size (>10,000) was enough for developing models. Indeed, some practitioners noted that systems based on data from a single or a few hospitals can be more beneficial because they would incorporate local trends of disease.

> *The sample size for the development of the model is not a problem for me. There are many more diagnostic rules that have less evidence than Emergency Alert System.* [EMR1]
>
> *I don't think the evidence should be considered weak just because the models are developed based on local data. Rather, I think it is beneficial because it could incorporate and reflect local trends of disease presentation.* [EP3]

Relative Advantage

Two informants recommended the system as a good alternative to diagnostic charts, in that it can reduce misdiagnoses. They implied Emergency Alert System showed superiority in that recollection is easier and it can help in pinpointing differential diagnoses, especially in atypical cases. Nevertheless, informants stated that the system would not be useful for typical cases in that the differential diagnoses would not change with the alert. Three informants added that the system can trigger a differential diagnosis for inexperienced physicians, and two informants argued that it may bias the clinician's decisions.

> *It is less likely to leave out diseases for experienced physicians. It is not necessary especially for physicians who have experience in the emergency department.* [EP2]
>
> *Easier to think of the differential diagnoses than a diagnostic chart.* [TYR3]
>
> *I think it would not be helpful for typical cases, though it can definitely be an aid for cases that are difficult to make the diagnosis.* [TYR3]
>
> *I feel that being alerted to a certain disease can cause bias, and it may lead to misdiagnoses.* [EP2]

Design Quality and Packaging

Ten informants talked about the design quality, and all of them considered the design to be neither interruptive for daily practice nor effortful and found it useful for medical practice. Moreover, the summary board made by the natural language processing algorithm was noted to be well visualized as a summary of information. Two informants mentioned the effect of real-time prediction and that the alert was intuitively visualized. However, one informant mentioned that if the system was extended to other diseases, the additional information could be complex and interruptive.

> *It is easy to see changes in the probability of the target disease, but I did not understand the timing of the alert.* [TYR6]
>
> *The summary screen was easy to understand.* [EMR2]
>
> *It is good that the alert comes out in real time.* [TYR2]
>
> *The system is simple, but I am concerned that alerts with many diseases would make it complex and interruptive.* [EMR3]

Compatibility

All informants noted that the system can be integrated with the existing workflow processes and practices. However, one informant stated that the system affected typing speed and could interfere with the clinical workflow if it were to stop working or freeze.

> *It is useful to obtain additional information without extra effort.* [EMR3]
>
> *Medical care can be done as usual. No difference from a regular electronic medical record.* [TYR5]
>
> *The system affects typing speed a bit. My biggest concern is that the system may interfere with the flow if the system freezes, or even worse, stops.* [EP1]

Discussion

Principal Findings

This study employed a mixed methods approach to analyze barriers to and acceptance of the implementation of AI-based CDSS. The quantitative analysis revealed generally positive responses for user acceptance and the qualitative analysis identified two significant barriers (System Performance, Compatibility) and two major facilitators (Evidence Strength, Design Quality) to implementation of AI-based CDSS. To our knowledge, this is the first study to analyze barriers, facilitators, and acceptance of AI-based CDSS implementation in an emergency care setting using a mixed methods approach.

User Acceptance

The quantitative analysis showed positive scores for Attitude and Intention to Use. Employing the UTAUT model, the achieved mean score of 3.63 indicated that if the Emergency Alert System were to be developed from the current prototype into a full software product, it would likely be well accepted by its users, as all scores were above the neutral score of 3, indicating favorable attitudes toward the use of the system. Nonetheless, since misdiagnoses could be life-threatening in the emergency setting, the barriers and facilitators identified in the qualitative analysis should be addressed thoroughly for the system to be well accepted.

Barriers

System Performance

The qualitative analysis implied that performance of the system could differ between typical and atypical cases, which could partly explain why the performance expectancy was not higher than expected from the quantitative analysis. Participants who had ≥3 years of clinical experience stated that alerts for typical cases would most likely be ignored and could moreover cause alert fatigue. There were also concerns that the clinicians' decisions could be biased by the system not only when the alert appears but also when it does not appear. As implied in a previous study [26], how we can prevent excessive trust in a CDSS, which can interfere with developing clinical skills in training physicians (eg, AI-based CDSSs focusing on ruling out critical conditions rather than making diagnoses), has yet to be fully explored. However, there were also positive comments that it can help improve accuracy of diagnosis and that it has the potential to decrease misdiagnosis—especially for inexperienced physicians—and reduce costs.

Compatibility

We found that the risk of system failure (eg, freezing of the system) is also a barrier, especially for AI-based CDSSs. In the interview, one participant reported previously experiencing system failure due to the installment of a new system, which disrupted the clinical workflow. Although some system failure issues have already been identified [27], the increasing computational resources required for AI should carefully be taken into consideration for both stand-alone CDSSs and CDSSs integrated with EHR. More importantly, providing physicians mental security that the system would not fail or freeze is crucial to user acceptance.

Facilitators

Evidence Strength

User distrust in the system has also been a barrier for CDSSs, and the "black box" nature of AI-based systems are known to compound this issue [28]. Through the explanation of the data set and by visualizing the alerts in an intuitive manner, the Emergency Alert System was perceived to be a sufficient alert system. In fact, some practitioners valued the locality over the universality of the system, considering the local trends of disease. Though this can be a facilitator to implementation, the validation of the model should be performed thoroughly, as small data sets tend to overfit, leading to undesired consequences [29].

Design Quality

Alert fatigue is a common issue in implementing CDSSs [30], and previous studies have revealed that alert fatigue could be reduced by refining the human-machine interface and clinical role tailoring [31]. Another study suggests that poor CDSS design could worsen alert fatigue [32]. The Emergency Alert System was designed to avoid alert fatigue through real-time background processing of information and by presenting visual information. The results of the quantitative analysis were consistent with the qualitative analysis in that users' perceived complexity of the system was low and they judged the system to be neither interruptive nor effortful. Although modifying the alert depending on the clinical role of the user (eg, nurses) is important to reducing alert fatigue, our results suggested that tailoring the system based on clinical experience would be more effective in practice.

Potential Limitations

There are potential limitations to this study. First, the participants reported that they have limited familiarity with information technology (although they do regularly use EHRs), which may mean that the findings might not be generalizable to other physicians. However, including physicians with varying years of postgraduate training from multiple tertiary care hospitals could have at least partially addressed this limitation. Second, the results may also be affected by external settings. The data were collected in two hospitals in Japan, the majority of clinicians who participated were trainees, and the developed model and user interface are unique to the CDSS. The convenience sampling and the composition of the clinical experience of participants in this study may have been sources of bias. Nevertheless, our findings are consistent with prior studies in that the user interface and system reliability are key to user acceptance and implementation [4,33,34]. Third, the sample size was small for a quantitative analysis and has limited real-world relevance. However, given the mixed methods design of our study, the sample size for the qualitative analysis is acceptable according to earlier studies [35,36]. Lastly, the system was in the preimplementation phase. Though most participants found the simulated setting to be comparable to the actual clinical setting, the session was done online and outside of the emergency department. Thus, further studies are needed to evaluate the developed system in the clinical setting. In addition, information regarding social influence and facilitating conditions were answered on the basis of respondents' personal knowledge and may not be consistent in the quantitative analysis.

Conclusions

Our mixed methods evaluation based on theoretically grounded frameworks revealed the acceptance, barriers, and facilitators of implementation of AI-based CDSS. Although the concern of system failure and overtrusting of the system could be barriers to implementation, the locality of the system and designing an intuitive user interface could likely facilitate the use of optimal AI-based CDSS. Alleviating and resolving these factors should be key to achieving good user acceptance of AI-based CDSS.

Acknowledgments

RF and TG conceived and designed the study. All authors interpreted the data, critically revised the paper for important intellectual content, and approved the final paper. RF, KH, and TS developed the Emergency Alert System. RF and KO performed the quantitative and qualitative analyses. KL and TG made clinical vignettes of model cases. NH and TO recruited interview participants. RF and TG drafted the initial paper. TG supervised the study. RF and TG are the guarantors. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Conflicts of Interest

The authors declare no competing nonfinancial interests but the following competing financial interests: RF, KL, SS, HN, and KH are paid researchers at TXP Medical Co Ltd, and TS and TG are the managers of TXP Medical Co Ltd.

Multimedia Appendix 1
Supplemental tables.
[DOCX File , 30 KB - formative_v6i6e36501_app1.docx]

References

1. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med 2020;3:17 [FREE Full text] [doi: 10.1038/s41746-020-0221-y] [Medline: 32047862]
2. Shimizu T, Nemoto T, Tokuda Y. Effectiveness of a clinical knowledge support system for reducing diagnostic errors in outpatient care in Japan: A retrospective study. Int J Med Inform 2018 Jan;109:1-4. [doi: 10.1016/j.ijmedinf.2017.09.010] [Medline: 29195700]
3. DeSilva MB, Kodet A, Walker PF. A Best Practice Alert for Identifying Hepatitis B-Infected Patients. Am J Trop Med Hyg 2020 Aug;103(2):884-886 [FREE Full text] [doi: 10.4269/ajtmh.20-0041] [Medline: 32431283]
4. Harry ML, Truitt AR, Saman DM, Henzler-Buckingham HA, Allen CI, Walton KM, et al. Barriers and facilitators to implementing cancer prevention clinical decision support in primary care: a qualitative study. BMC Health Serv Res 2019 Jul 31;19(1):534 [FREE Full text] [doi: 10.1186/s12913-019-4326-4] [Medline: 31366355]
5. Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nat Med 2019 Jan;25(1):65-69 [FREE Full text] [doi: 10.1038/s41591-018-0268-3] [Medline: 30617320]
6. Combi C, Pozzi G. Clinical Information Systems and Artificial Intelligence: Recent Research Trends. Yearb Med Inform 2019 Aug;28(1):83-94 [FREE Full text] [doi: 10.1055/s-0039-1677915] [Medline: 31419820]
7. Richardson S, Cohen S, Khan S, Zhang M, Qiu G, Oppenheim MI, et al. Higher Imaging Yield When Clinical Decision Support Is Used. J Am Coll Radiol 2020 Apr;17(4):496-503 [FREE Full text] [doi: 10.1016/j.jacr.2019.11.021] [Medline: 31899178]

8. Long D, Capan M, Mascioli S, Weldon D, Arnold R, Miller K. Evaluation of User-Interface Alert Displays for Clinical Decision Support Systems for Sepsis. Crit Care Nurse 2018 Aug;38(4):46-54 [FREE Full text] [doi: 10.4037/ccn2018352] [Medline: 30068720]
9. Khairat S, Marc D, Crosby W, Al Sanousi A. Reasons For Physicians Not Adopting Clinical Decision Support Systems: Critical Analysis. JMIR Med Inform 2018 Apr 18;6(2):e24 [FREE Full text] [doi: 10.2196/medinform.8912] [Medline: 29669706]
10. Segal MM, Rahm AK, Hulse NC, Wood G, Williams JL, Feldman L, et al. Experience with Integrating Diagnostic Decision Support Software with Electronic Health Records: Benefits versus Risks of Information Sharing. EGEMS (Wash DC) 2017 Dec 06;5(1):23 [FREE Full text] [doi: 10.5334/egems.244] [Medline: 29930964]
11. Berner ES. Diagnostic decision support systems: why aren't they used more and what can we do about it? AMIA Annu Symp Proc 2006:1167-1168 [FREE Full text] [Medline: 17238783]
12. Venkatesh V, Morris MG, Davis GB, Davis FD. User Acceptance of Information Technology: Toward a Unified View. MIS Quarterly 2003;27(3):425. [doi: 10.2307/30036540]
13. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 07;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
14. MAXQDA. 2019. URL: http://www.maxqda.com [accessed 2021-10-01]
15. Criado FJ. Aortic dissection: a 250-year perspective. Tex Heart Inst J 2011;38(6):694-700 [FREE Full text] [Medline: 22199439]
16. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. 2016 Aug 13 Presented at: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 13-17, 2016; San Francisco, CA, USA. [doi: 10.1145/2939672.2939785]
17. Rodríguez-Pérez R, Bajorath J. Interpretation of machine learning models using shapley values: application to compound potency and multi-target activity predictions. J Comput Aided Mol Des 2020 Oct;34(10):1013-1026 [FREE Full text] [doi: 10.1007/s10822-020-00314-0] [Medline: 32361862]
18. Venkatesh V, Sykes T, Zhang X. 'Just What the Doctor Ordered': A Revised UTAUT for EMR System Adoption and Use by Doctors. 2011 Jan 4 Presented at: 2011 44th Hawaii International Conference on System Sciences; January 4-7, 2011; Hawaii. [doi: 10.1109/HICSS.2011.1]
19. Venkatesh V, Thong J, Xu X. Unified Theory of Acceptance and Use of Technology: A Synthesis and the Road Ahead. JAIS 2016 May;17(5):328-376. [doi: 10.17705/1jais.00428]
20. Tavakol M, Dennick R. Making sense of Cronbach's alpha. Int J Med Educ 2011 Jun 27;2:53-55 [FREE Full text] [doi: 10.5116/ijme.4dfb.8dfd] [Medline: 28029643]
21. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
22. Consolidated Framework for Implementation Research (CFIR). URL: https://cfirguide.org [accessed 2022-05-23]
23. Srivastava A, Thomson S. Framework Analysis: A Qualitative Methodology for Applied Policy Research. Journal of Administration and Governance 2009:72-79.
24. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Med Res Methodol 2013 Sep 18;13:117 [FREE Full text] [doi: 10.1186/1471-2288-13-117] [Medline: 24047204]
25. Ritchie J, Spencer L. Qualitative data analysis for applied policy research. In: Analyzing Qualitative Data. London: Routledge; 2002.
26. Ash JS, Sittig DF, Campbell EM, Guappone KP, Dykstra RH. Some unintended consequences of clinical decision support systems. AMIA Annu Symp Proc 2007 Oct 11:26-30 [FREE Full text] [Medline: 18693791]
27. Wright A, Hickman TT, McEvoy D, Aaron S, Ai A, Andersen JM, et al. Analysis of clinical decision support system malfunctions: a case series and survey. J Am Med Inform Assoc 2016 Nov;23(6):1068-1076 [FREE Full text] [doi: 10.1093/jamia/ocw005] [Medline: 27026616]
28. Asan O, Bayrak AE, Choudhury A. Artificial Intelligence and Human Trust in Healthcare: Focus on Clinicians. J Med Internet Res 2020 Jun 19;22(6):e15154 [FREE Full text] [doi: 10.2196/15154] [Medline: 32558657]
29. Deo RC. Machine Learning in Medicine. Circulation 2015 Nov 17;132(20):1920-1930 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.115.001593] [Medline: 26572668]
30. Ancker JS, Edwards A, Nosal S, Hauser D, Mauer E, Kaushal R, with the HITEC Investigators. Effects of workload, work complexity, and repeated alerts on alert fatigue in a clinical decision support system. BMC Med Inform Decis Mak 2017 Apr 10;17(1):36 [FREE Full text] [doi: 10.1186/s12911-017-0430-8] [Medline: 28395667]
31. Hussain MI, Reynolds TL, Zheng K. Medication safety alert fatigue may be reduced via interaction design and clinical role tailoring: a systematic review. J Am Med Inform Assoc 2019 Oct 01;26(10):1141-1149 [FREE Full text] [doi: 10.1093/jamia/ocz095] [Medline: 31206159]

32. Yang Q, Steinfeld A, Zimmerman J. Unremarkable AI: Fiting intelligent decision support into critical, clinical decision-making processes. 2019 Presented at: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow, Scotland, UK. [doi: 10.1145/3290605.3300468]
33. Arts DL, Medlock SK, van Weert HCPM, Wyatt JC, Abu-Hanna A. Acceptance and barriers pertaining to a general practice decision support system for multiple clinical conditions: A mixed methods evaluation. PLoS One 2018;13(4):e0193187 [FREE Full text] [doi: 10.1371/journal.pone.0193187] [Medline: 29672521]
34. Jung SY, Hwang H, Lee K, Lee H, Kim E, Kim M, et al. Barriers and Facilitators to Implementation of Medication Decision Support Systems in Electronic Medical Records: Mixed Methods Approach Based on Structural Equation Modeling and Qualitative Analysis. JMIR Med Inform 2020 Jul 22;8(7):e18758 [FREE Full text] [doi: 10.2196/18758] [Medline: 32706717]
35. Guest G, Bunce A, Johnson L. How Many Interviews Are Enough? Field Methods 2016 Jul 21;18(1):59-82. [doi: 10.1177/1525822X05279903]
36. Braun V, Clarke V. Successful Qualitative Research: A Practical Guide for Beginners. London: Sage; 2013.

Abbreviations

AI: artificial intelligence
CDSS: clinical decision support system
CFIR: Consolidated Framework for Implementation Research
COREQ: Consolidated Criteria for Reporting Qualitative Research
EHR: electronic health record
UTAUT: Unified Theory of Acceptance and Use of Technology

Edited by A Mavragani; submitted 16.01.22; peer-reviewed by K Trinkley, H Yun, G Hackbarth; comments to author 21.02.22; revised version received 13.04.22; accepted 08.05.22; published 13.06.22.

Please cite as:
Fujimori R, Liu K, Soeno S, Naraba H, Ogura K, Hara K, Sonoo T, Ogura T, Nakamura K, Goto T
Acceptance, Barriers, and Facilitators to Implementing Artificial Intelligence–Based Decision Support Systems in Emergency Departments: Quantitative and Qualitative Evaluation
JMIR Form Res 2022;6(6):e36501
URL: https://formative.jmir.org/2022/6/e36501
doi:10.2196/36501
PMID:35699995

Original Paper

Interest in HIV Prevention Mobile Phone Apps: Focus Group Study With Sexual and Gender Minority Persons Living in the Rural Southern United States

Jeb Jones[1], MPH, MS, PhD; O Winslow Edwards[1], MPH; Leland Merrill[2], BA; Patrick S Sullivan[1], DVM, PhD; Rob Stephenson[2], PhD

[1]Department of Epidemiology, Emory University, Atlanta, GA, United States

[2]Department of Systems, Populations and Leadership, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:
Jeb Jones, MPH, MS, PhD
Department of Epidemiology, Emory University
1518 Clifton Road NE
Atlanta, GA, 30322
United States
Phone: 1 4047122275
Email: jeb.jones@emory.edu

Abstract

Background: Mobile health (mHealth) interventions, including smartphone apps, have been found to be an effective means of increasing the uptake of HIV prevention tools, including HIV and sexually transmitted infection (STI) tests and pre-exposure prophylaxis. However, most HIV prevention mHealth apps tested in the United States have been tested among populations living in areas surrounding urban centers. Owing to reduced access to broadband internet and reliable cellular data services, it remains unclear how accessible and effective these interventions will be in rural areas. In addition, gay and bisexual men who have sex with men and gender minority populations in rural areas experience enhanced stigma when compared with their more urban counterparts, and these experiences might affect their willingness and interest in mHealth apps.

Objective: This study aimed to conduct online focus groups with men who have sex with men and transgender and gender diverse populations in the rural southern United States to assess their interest in mHealth HIV prevention apps and the features that they would be the most interested in using.

Methods: Focus group participants were recruited from a larger pool of sexual and gender minority respondents to a web-based research survey. The participants indicated that they would be willing to participate in an online focus group discussion. Focus groups were conducted via secure Zoom (Zoom Video Communications Inc) videoconferencing. During the focus group discussions, participants were asked to discuss their experiences with HIV and STI prevention and how these experiences were affected by living in a rural area. They were then shown screenshots of a new app to promote HIV and STI prevention among rural populations and asked to provide their opinions on the app's features. The transcripts of the discussions were reviewed and coded using a constant comparative approach.

Results: A total of 6 focus groups were conducted with 26 participants. Most participants were cisgender gay and bisexual men who have sex with men (19/26, 73%); the remaining participants were transgender men (2/26, 8%), were nonbinary people (2/26, 8%), or had multiple gender identities (3/26, 12%). Participants reported numerous barriers to accessing HIV and STI prevention services and accurate information about HIV and STI prevention options. Overall, the participants reported a high degree of interest in mHealth interventions for HIV and STI prevention and suggested several recommendations for the features of an app-based intervention that would be the most useful for rural residents.

Conclusions: These focus group discussions indicate that rural residence is not a major barrier to mHealth HIV and STI prevention intervention implementation and that there is a high degree of interest in these approaches to HIV and STI prevention.

(JMIR Form Res 2022;6(6):e38075) doi:10.2196/38075

KEYWORDS

men who have sex with men; transgender persons; nonbinary persons; mHealth; mobile app; HIV; pre-exposure prophylaxis; PrEP; sexually transmitted infection testing; STI testing; HIV testing; mobile phone

Introduction

Background

The Ending the HIV Epidemic (EHE) initiative has identified 57 jurisdictions for increased HIV prevention resources [1]. Although many of these jurisdictions encompass urban centers, 7 states with large rural populations are identified as priority jurisdictions. Of these 7 states, 6 (Alabama, Arkansas, Kentucky, Mississippi, Oklahoma, and South Carolina) are in the southern United States and the seventh (Missouri) is geographically contiguous with the southern United States. These states were targeted because at least 10% of new HIV diagnoses in each of these states occurred in rural areas in 2016 and 2017. Similar to urban areas of the United States, gay and bisexual men who have sex with men (GBMSM) account for most of the new diagnoses in rural areas [2]. Despite the importance of rural communities in the EHE initiative, research is lacking on the sexual behavior and health care preferences of sexual and gender minority populations in rural areas.

Rural GBMSM are less likely to engage in important HIV prevention behaviors, including HIV and sexually transmitted infection (STI) testing and pre-exposure prophylaxis (PrEP) use [3,4]. They face increased and context-specific barriers to accessing HIV and STI prevention resources compared with men in more urbanized areas [5-7]. On the supply side, rural GBMSM are less likely to have access to culturally competent care [5-8], and PrEP awareness and lack of comfort in prescribing PrEP among providers have been barriers to PrEP uptake among rural GBMSM [7]. On the demand side, stigma and discrimination are barriers to engaging in HIV prevention. GBMSM in rural areas are less likely to disclose their sexual identity to their health care provider [9], report being afraid to seek health care, and avoid health care settings more frequently than urban GBMSM [10]. Experiences of stigma are heightened among rural GBMSM owing to more conservative attitudes toward same-sex sexual behavior and more insular social environments [11]. Intersecting minoritized identities intensifies these experiences among GBMSM of color [12].

Fewer HIV prevention studies have been conducted among transgender people living in rural areas; however, it is likely that the barriers to health care access experienced by transgender people in general [13,14] are exacerbated among transgender people living in rural communities. One study that included transgender women living in rural Florida found that transgender women were more likely to receive a late HIV diagnosis than their cisgender counterparts [15]. Transgender men in rural areas have been found to be less likely to have a primary care provider or to have had a blood cholesterol screening compared with urban transgender men [16].

Mobile health interventions (ie, apps) offer a potential solution to break down some of the barriers to HIV prevention services for rural sexual and gender minority individuals. Although data are not available for sexual and gender minority populations specifically, approximately 80% of rural adults use smartphones and 66% use social media [17,18]. In previous studies, GBMSM in rural areas have indicated that telehealth solutions for HIV and STI screening are acceptable methods for receiving these services [19,20]. Smartphone apps provide a discreet means of delivering sexual health information and can be a platform for delivering telehealth services, including HIV and STI testing and PrEP. Apps have been found to be acceptable to GBMSM [21], and a number of ongoing studies have assessed the efficacy of apps to increase HIV prevention services uptake among this priority population [22-26]. One app has shown efficacy in increasing the uptake of HIV testing and PrEP among higher-risk GBMSM [27]. However, studies testing the feasibility or efficacy of apps have tended to enroll populations recruited from urban centers or periurban areas. Men living in rural areas might hold heightened concerns regarding privacy and disclosure of sexual behavior or gender identity; these concerns might affect their willingness to use these apps [28]. Owing to these additional barriers faced by GBMSM and transgender people in rural areas, coupled with reduced limited access to high-speed internet and cellular services in some areas, data are needed on the feasibility and acceptability of app-based HIV prevention interventions for GBMSM and transgender people in rural areas.

Objectives

We conducted focus group discussions with GBMSM and transgender and gender nonconforming people in the rural southern United States to assess health care use and willingness to use a mobile app to access sexual health information and order HIV or STI test kits to assess the feasibility of implementing app-based HIV prevention interventions in the rural southern United States.

Methods

Participants

Participants were recruited using web-based advertisements on Facebook, Instagram, Jack'd, Grindr, and Reddit and on social media feeds of community-based organizations serving rural sexual and gender minority communities. After completing a web-based eligibility screener, the participants completed a web-based survey and indicated their willingness to participate in an online focus group discussion. The eligibility screener and survey were only available in English. Eligible participants were aged 18 to 34 years; were cisgender men, transgender men, transgender women, and nonbinary people; reported ever having anal or vaginal sex; self-reported HIV negative; owned an iPhone or Android smartphone; and lived in a rural area of the southern United States, as defined by the US Census Bureau or the state of Missouri. Eligible participants were purposively sampled to obtain a diverse group with respect to age and race and ethnicity. More than half of all new HIV diagnoses in the United States occur in the South [29]. Missouri is geographically contiguous with the South and is a priority jurisdiction for the EHE campaign [1] because of the high burden of new HIV

diagnoses occurring in rural areas of the state. Participants were compensated US $50 via an electronic gift card following the completion of the focus group discussion.

Definition of *Rural*

Participants reported their ZIP code of residence, which was then mapped to the county of residence using an established algorithm [30]. Rural counties were those that were classified as micropolitan or noncore by the National Center for Health Statistics [31], had a Rural-Urban Commuting Area Code of ≥4 [32], or had an Index of Relative Rurality score of ≥0.4 [33]. None of these methods is designed specifically to categorize communities with respect to the availability of culturally competent care for GBMSM or transgender individuals, and it is not clear which of these schemes is most relevant to differentiating between rural and nonrural communities for this purpose. Thus, we used multiple definitions to have an inclusive criterion for participant eligibility.

App

The Combine app is an adaptation of HealthMindr (Emory University in collaboration with Keymind and Softura) [21], a comprehensive HIV prevention app originally developed for cisgender GBMSM. The Combine app has the following functionality: frequently asked questions about HIV and STIs, PrEP, postexposure prophylaxis, and health insurance; recommendations for testing frequency; the ability to order HIV or STI self-test kits, condoms, and condom-compatible lubricants; behavioral risk self-assessments; and provider locators for PrEP and HIV or STI testing.

Focus Groups

Focus groups were conducted on the web using Zoom (Zoom Video Communications Inc) and were recorded for transcription. Participants had the option of using a nickname and leaving their cameras off for anonymity. The focus groups started with a discussion about where participants usually accessed HIV and STI testing, experience with and interest in accessing HIV and STI testing on the web, experiences accessing health care and any issues encountered based on sexual or gender identity, and how living in a rural area affected their willingness and ability to access sexual health care services. Next, the participants were asked to describe the features of the smartphone apps that they liked and disliked. The participants were then asked to describe where they found information about sexual health (eg, on the web, from friends, or from health care providers). Finally, the participants were shown screenshots of the Combine app. After viewing the screenshots, the participants were asked to provide feedback on the app including interest in using the individual sections; willingness to order HIV and STI self-test kits, condoms, and lubricants through the app; whether they thought their peers would use it; and any functions that might be missing. Focus groups were stratified by gender identity so that cisgender men were grouped together and transgender persons, nonbinary people, and those with other gender identities were grouped together. Focus groups were conducted until saturation occurred overall but not necessarily within subgroups of gender identity (ie, no additional novel information was being generated). All the focus groups were facilitated by one of the coauthors (LM).

Analysis

Transcriptions of focus group discussions were coded by 2 coders (LM and OWE) using a constant comparative approach [34]. Using this approach, coders first read the transcripts and identified broader emergent codes, which were used to construct an initial codebook. After an initial review of the multiple transcripts, the coders came together to discuss and clarify the meaning of each code. Subsequently, codes were applied to all transcripts, and newly emergent codes were added to the codebook. After this full pass, the coders once again met to probe the meanings of codes and to define higher-level themes that were emerging. In addition to the emergent themes that were identified, the a priori themes of anticipated and enacted stigma were included. Following this meeting, all transcripts were coded a second time using the finalized codebook.

Ethics Approval

The participants provided informed consent to participate in focus group discussions. All study procedures were reviewed and approved by the Emory University Institutional Review Board (protocol 00001268).

Results

Participant Demographics

A total of 91 participants (77 cisgender men, 8 nonbinary people, 2 transgender men, and 4 with multiple gender identities) were eligible and expressed willingness to participate in the focus groups. Of these 91 participants, 26 (29%) ultimately agreed to participate and contributed to 6 focus groups, comprising 2 to 8 participants per group. The demographic characteristics of the participants are presented in Table 1. Most participants (19/26, 73%) were cisgender GBMSM; the remaining participants were transgender men (2/26, 8%), nonbinary people (2/26, 8%), or had multiple gender identities (3/26, 12%).

Table 1. Demographic characteristics of focus group participants (N=26).

Demographics	Value
Age (years), median (IQR)	25 (21-29)
Race or ethnicity, n (%)	
Asian participants	1 (4)
Black participants	6 (23)
Mixed race participants	18 (69)
White participants	1 (4)
Hispanic participants, n (%)	
Yes	1 (4)
No	25 (96)
Sex assigned at birth, n (%)	
Male	22 (85)
Female	3 (12)
Intersex	1 (4)
Gender identity, n (%)	
Cisgender men	19 (73)
Transgender men or transmasculine	2 (8)
Nonbinary or gender nonconforming	2 (8)
Multiple identities	3 (12)
Education, n (%)	
High school or GED[a]	3 (12)
Some college	10 (38)
College graduate	13 (50)
Insurance, n (%)	
Private	20 (77)
Public	2 (8)
Other	2 (8)
None	2 (8)
Annual income (US $), n (%)	
<19,999	3 (12)
20,000 to 39,999	5 (19)
40,000 to 74,999	7 (27)
>75,000	7 (27)
Prefer not to answer or do not know	4 (15)

[a]GED: General Educational Development test (an alternative to a high school diploma in the United States).

Focus Group Discussions

Qualitative data analysis identified 5 major themes, which are described in the next sections, with representative quotations from participants.

Access to Health Services

Participants talked about access to health services both in terms of their ability to access health care directly and in terms of access to transportation and internet services to facilitate the uptake of health care services. Participants often reported that accessing lesbian, gay, bisexual, transgender, and queer+ (LGBTQ+)–competent health care was not always possible in the smaller towns where they lived:

> *I'd say I'd probably have to travel about an hour to the nearest big city that's accepting...'Cause my area is not exactly the most accepting, in that regard.*

Others have described issues where STI testing might only be available on certain days in their community if not ordered

through a primary care provider. In contrast, some participants described being able to access sexual health care in their jurisdiction from an LGBTQ+-specific health center, a primary care provider, or some other sexual health provider:

> *Luckily, where I live, there's kind of a specific clinic for...I think it used to be called [name of organization], but they've kind of transitioned into a more comprehensive HIV prevention and care place. So they are, I think, probably the main providers of PrEP and STI services here.*

For those who did not have an accepting clinic in close proximity, access to transportation could facilitate or limit the use of LGBTQ+-competent health services in larger cities:

> *Well, at least for me, because of my disability, transportation issues is a major factor, as well. There are no Planned Parenthoods in my area, and the only health organization that would otherwise do STI testing is through the closest university, which is still like 25 minutes away, so it's not very fast.*

Another participant noted that there were providers available locally, but the fear of being outed made them undesirable. Thus, transportation was a barrier despite proximity to the available providers:

> *When I was younger, it was more of a barrier, because it's a small town, so all of my...All the providers kind of knew me and my family, so that was definitely...And no way of travel, was a barrier...Something that was very anxiety-provoking, when I needed...I knew that I maybe needed a test, but then going to access one was kinda traumatic, so I avoided it sometimes.*

Cultural Environment

Cultural environment described the experience of living in small, rural communities wherein privacy is limited, both inside and outside of the household, and socially conservative values are prevalent. Participants consistently described town dynamics where community members were curious about one another's private lives and privacy was difficult to maintain. To avoid gossip, this participant chose to use at-home testing instead of seeking sexual health care in person:

> *I go to a Christian college, so it's kind of difficult around the area to go in for screenings, just because it's kind of a small town as well, everybody talks, so it's not very welcomed.*

As demonstrated by the following participant, a common concern was that one's sexuality could be revealed to the community as a result of using sexual health services or products:

> *Well, because when you have to go to the store for those things, you gotta go in person and so...I don't know a whole lot about straight sex, but I'm sure they're not all buying lube and...So you went and bought lube and now everybody knows that you're [chuckle] doing something.*

In addition to this lack of privacy, participants described navigating socially conservative values and stigma toward LGBTQ+ people, both in the community and in health care contexts. A major concern was the lack of relevant sexual health knowledge, which participants traced back to inadequate and stigmatizing sexual education experiences in school:

> *Growing up, I went to a religious high school and sex education was nonexistent and my parents didn't really teach me anything. In my biology class, I mean it said things like "erection" and "clitoris" and things like that. The school actually glued the pages together in our textbook, so we couldn't see them.*

Conservative values and stigmatizing views were also encountered in medical settings. Many participants described experiences where staff and providers were tangibly uncomfortable working with them:

> *Honestly, here where I'm from, Arkansas, they're just...Everything is kinda backwards in here. It's like as soon as they figure out that you're of the LGBTQ community, they almost have like a step back, "Oh my God." It's weird.*
>
> *I think I face more stigma because I'm in the sex work industry, and is going to a PCP is definitely hard to explain why I need to get tested so often. And then there's just that stigma surrounding that, it just makes it incredibly uncomfortable.*

Other participants described anticipating stigma when seeking medical care:

> *Well, I just moved, like I said, and I haven't yet found a PCP. And so, I would have to get over that barrier once more, like fear of judgment from a PCP, so I'm trying to find a new one, and right now I'm going to Planned Parenthood.*

As the following participant explains, although doctors are expected to be professional and exercise confidentiality, they are also entrenched in the community and sociopolitical milieu in which they live and work:

> *I had a primary care physician that I shared with my immediate family for...'Cause I'm from a very small town so there's not a lot of options. And once I came out, it just felt a little bit different 'cause I know...It's very small so everybody knows everybody and people are doctors, but they're also just people in your community as well. So, once I actually went to college and graduated out of my parents' insurance and got my own insurance, I didn't really feel the need to continue to stay at that same place.*

Discretion or Confidentiality

Participants in all the focus groups frequently emphasized the necessity of *discretion* in the form of privacy and confidentiality when it came to seeking or accessing sexual health services. These concerns were particularly heightened because of the lack of privacy and conservative values implicit in the *cultural environment* and the potential for shared familial housing arrangements:

> *But yes, especially a sense of discretion, if we're talking about queer people in rural areas. I know my hometown, my one-stoplight hometown, if anybody found out it would be a huge shunning issue in the town.*

For participants who may not be out to their local community or family, confidentiality is not only necessary to maintain their safety and social standing but also a prerequisite for their willingness to access sexual health services. Participants needed to know that their information would be kept confidential and that discretion would be prioritized, whether by local health care providers or by an app. One participant relayed an experience where confidentiality was compromised:

> *And while I consider what happened as coincidental or incidental, [a clinic employee] actually outed me and that really changed the dynamic of my whole life...It's sort of like having a conversation, "Hey [name], I saw...Oh yeah, you came down to XYZ Clinic." And everyone knows that clinic to be a specific clinic for the LGBTQ+...And I was like, "Oh." And you can probably follow the line from there.*

Owing to concerns about discretion when accessing care in person, some participants preferred to use telehealth and home testing offered by services such as Mistr, a web-based sexual health provider that caters to the LGBTQ+ community [35]. The participants identified the ability to receive confidential testing and care from the privacy of their own space as a major benefit of telehealth.

When discussing apps, the participants felt that discretion should be implemented at all levels, from the visual branding of the app to the app's privacy policies to the need to identify themselves within the app:

> *I know certain apps that used levels of discretion will name it something else or even have the ability to change what icon it is. I've seen things where you don't necessarily want someone to know you have an app of X, Y or Z, so it'll show up as a calculator or something like that on your phone. If somebody is looking over your shoulder and you're thumbing through your apps, if you have someone nosy, they don't see something specific for it, it just looks like another calculator app, or mix up possibly the ability to change the icon or something along that line, or maybe an abbreviation that only a few people would understand to maybe help with a level of discretion.*

One participant suggested that the app be password protected, which would help youth and others whose phones may be checked by family members keep their information private. When it came to marketing the app and the mailing of sexual health tests or supplies, discreet packaging was recommended to help maintain the privacy of the recipient:

> *I think the only thing I could really think of is just making sure that the materials were sent in a very discreet box. I live in a very small town and everyone knows each other, including the mailman, so...I think it being discreet is critical.*
>
> *...If you're living with your parents, maybe not the most great thing for them to see on their doorstep. But if you are living alone or living elsewhere, I think, at that point, it's okay with discretion. I think it's just better to be safe and more discreet than sorry, in a way.*

Packaging that maintained the recipient's privacy was very important, especially for those who might be living in multigenerational households. Finally, for transgender and nonbinary participants, privacy was also extended to the name used in the packaging. In some housing situations, transgender and nonbinary participants may need to use their birth name for safety, whereas in other contexts, they may be able to use their chosen name. Having the option to specify which name to use when ordering sexual health tests or supplies was a way of maintaining *discretion* around one's gender identity:

> *I also think the name thing, when you go to order an item, you can do your preferred name or the legal name or whatever, and have that option right in front of you. Because if you're living one place, one time, and you can use your preferred name when you want your orders with that, or if you're living in another place, you need it as the other name, whatever that may be.*

Convenience

In rural settings where culturally competent health care providers may be located far away and transportation may be inaccessible, *convenience* was paramount to accessing sexual health services and materials. In the context of sexual health services and supplies, *convenience* referred not only to ease of use but also to timeliness, affordability, and ease of attainment. A participant described the convenience of home testing compared with visiting a provider:

> *I do like the ease and convenience of using a service like Mistr. And they do have a referral program, and I've sent it to a few of my friends just because I'm getting [PrEP] for free, and it's very easy. It's delivered to me. I don't have to go to Walgreens and have them tell me that they don't have my prescription that day, 'cause it's on back order or whatever...I like the immediacy.*

When sexual health care was not convenient, whether it took too much time to access or was hard to afford or attain, participants expressed that they might be more likely to put it off:

> *I was looking because I didn't want to have to travel and take time out of the day to go make an appointment and find somewhere. Where I live, you can only do STI testing on certain days if you don't go to your PCP. I just thought it'd be more convenient to do it at home, but my insurance wouldn't cover it, and it seems like the cost, I was in law school at the time, it was somewhat cost-prohibitive, so I just ended up submitting late.*

Participants expressed eagerness to use sexual health care options they deemed convenient, and this followed through to their interest in the app. Considering the app, many participants appreciated that it was a *one-stop shop* for HIV testing, sexual health products, mental health screenings, and locating providers:

> *I think that it's gonna be potentially super good for...Especially for people in rural areas, not just the convenience and the discretion, but just even having the access to those services. And the convenience of an app and having it all in one area is awesome.*

When speaking about the locators in the app that allow users to find HIV or STI testing, PrEP providers, and mental health and substance use treatment providers in their area, users thought that these features would be useful in identifying conveniently located providers:

> *I love this idea [locator], mainly because it's convenient, not a lot of people know where they can get PrEP and information on PrEP or things like that, so this is really convenient.*

Participants were also very interested in low-cost or free services through the app, given that affordability and insurance status could be barriers:

> *I think if there's funding for some of those resources to be mailed out for free, that's probably the most, I would say, really beneficial aspect.*

Another participant shared how a sexual health app would benefit him:

> *Right now, I don't have a means of transportation, and so having something like this [app], where I can have things delivered or I can chat with a professional over the phone, I find that really convenient.*

Thus, participants viewed the app favorably and perceived it to reduce various barriers to care.

Trust or Comfort

Trust and *comfort* came up in 2 primary contexts: when ascertaining the quality of health information and when considering one's relationship with a health care provider. When it came to information, participants wanted content that they deemed to be accurate, legitimate, and trustworthy. Many participants were very discerning when evaluating health information and checked multiple sources to ensure that information was factual:

> *I'd like to check out a few different sources just because I want to get the most accurate picture of today's standard practices, and if anything new has been found out...Yeah, I've looked at WebMD, Healthline...Planned Parenthood's website is a good resource. Yeah. Usually those...Also, I like to read white papers. If I need to find something out, I'll look it up on PubMed or something else too.*

Often, the origins of health information were considered when determining whether content was factual; that is, participants were more inclined to believe health information that came from a trusted friend or website than information they might find through a search engine or social media:

> *If it's coming from the Kaiser Family Foundation, I know that I can trust it just because they have epidemiologists and PhDs and experts in public health that are providing that information, but if it comes from Fox News and it's not cited, questionable.*

> *I have a few friends that if I have any kind of need for any kind of...Not just necessarily to search for help, but in general. I can just be like, "Hey, do y'all have any experience with this?" and if they tell me no, then I'll probably go to Google or just ignore it until it becomes a bigger problem.*

If a source was trusted, a participant would feel *comfortable* or willing to use that source. When it came to providers, *trust* was also intertwined with *comfort*, an understanding that a provider would be accepting of one's identity, provide or foster a feeling of safety, and be able to provide LGBTQ+-competent care:

> *Honestly, I'd feel really comfortable with finding another LGBT resource center and something specifically for people like me. So, if...Not really like Planned Parenthood, but something like that, I'd feel really comfortable in an environment like that.*

For participants, it was important to locate a provider who could offer caring, LGBTQ+-competent services. On top of that, building a close working relationship with a provider was conducive to high-quality care. On the flip side, some participants struggled to locate providers that were accepting, which was partially attributed to the issues of *cultural environment* and stigma:

> *I think for me, I get the primary care physician being the...I can see why that's more desirable. I think for me, I just don't know how to go about finding somebody that I would feel comfortable with...Especially, I think maybe as [other participant] mentioned, like small towns and stuff like that, that stuff's really hard there, I think.*

This participant also identified moving frequently and aging out of pediatric care as contributors to the struggle to find a trusted primary care provider. As a student who was about to graduate, they were not sure what they would do for sexual health services upon graduating. Other participants were able to access LGBTQ+-specific providers but felt pigeonholed and uncomfortable with their approach to treatment:

> *Well, going there, going to my primary care physician, they don't see me as a gay, bisexual male. But when I go to a health center, they treat me as...It's sort of like, "Oh, a gay person is here so we're gonna do this." And the type of service that I receive from them is completely different than the type of service and care that I receive from my primary care physician where I use my insurance.*

For this participant, it was important to be viewed holistically and not simply through the lens of his sexuality. Therefore, he had switched from a community health center to a primary care provider. When participants did not feel *trust* or *comfort* in a

provider, locating and using health services were major concerns:

> *I think that one of the things that, before I even use the app, we were talking about credibility earlier, and I think that I would trust the information on this app, if either there were sources or when you open the app, it's like it says the name of the app, but it's like powered by or created by Emory or some university. I feel like that would give it some credence.*

Preferences for a Mobile HIV and STI Prevention App

Overview

App features describe feedback directly pertaining to what participants were looking for when accessing sexual health information and care in a digital environment. Many of these responses were elicited in response to discussion of mock-ups from the Combine app or in discussion of other apps and web-based resources that participants use frequently. Important features included *relevant information* and *functional recommendations*.

Relevant Information

Participants expressed a desire for information that was relevant to their experiences as members of the LGBTQ+ community when searching for sexual health information, whether in person, on the web, or through a health app. It was important for participants to feel that the health information and care they received were tailored and *relevant* to their needs as an LGBTQ+ person.

Given that many participants discussed receiving inadequate sexual education, they were very interested in having a detailed frequently asked questions section that would provide accessible information about options for safe sex and navigating consent, including introductory topics:

> *I don't know exactly how this could fit into the app, but when talking about sexual health, just...Yeah, even like the condom use page, maybe if there was even like a diagram, I mean even with something like a dildo or whatever, "This is how you put a condom on." Because I never was shown how to use one. I haven't had a ton of sexual partners, but I've actually never used one, and so that would have been really nice to know or other options for that.*

Another participant added:

> *In addition, to just safe sex practices, also just things like healthy relationships and consent, and that sort of stuff, I think that could be really beneficial, especially for young people in rural areas....*

For transgender, nonbinary, and bisexual or pansexual participants, it was important to have information that was applicable and affirming to their gender identity and anatomy and the anatomy of their sexual partners:

> *Yeah, I think it was on Gilead's website, or maybe it's on aids.gov that they ask you what your gender identity is, what your anatomy is, and then they ask you a question about the anatomy and gender identity of your partners. So, that way they can customize the suggestions for you, so I don't know...That would be like the first thing when you create a profile at the first time you open the app, and then if you need to change it, you can hit the wrench or the gear and change it another time.*

Ultimately, participants were ready to engage with relevant, tailored information about sexual health.

Functional Recommendations for HIV and STI Prevention Mobile Apps

Participants provided various *functional recommendations* about features that could be added to the app to improve its usability and better serve users' needs. Many participants wanted the app to integrate easily with existing utility apps on their phone. For example, one feature of the Combine app is a timeline that displays study milestones and planned prevention activities (eg, scheduled HIV tests and reminders to order condoms). One participant noted that they would want to integrate this feature with other reminders on their phone:

> *I really like the timeline thing, which is I know that if I don't have things in front of me, they don't exist. I think it would also be a really useful feature to have a calendar integration, add the dates to your calendar. But I also really like how you can see all of the different steps that you have to participate in. That was a feature that I really appreciated.*

Participants were interested in the ability to track shipments of sexual health materials or testing kits ordered via the app:

> *I think that there's something about being able to monitor and track something in real-time is very gratifying, and if you can ever...I feel like I ever have a service that allows me to do that, it kinda makes me wanna use it more. One of the reasons I have shipped through FedEx is because their online tracking is so solid. I think just having the ability to kinda track and feel like you're in constant touch with whatever it is you're trying to do is a really nice thing to have.*

Participants wanted the ability to easily tailor their search to locate content within the app, whether sexual health information or providers that offer specific services:

> *I think of it like Zillow, where if you wanna just look at properties, you can look at properties, but then you can filter based on like condo or house. I like that it has a list of options in the geographic location. I don't know if every center has testing, screening, if hopefully, they're all queer and trans-friendly. But I don't know that, I'm just assuming things like the MinuteClinic is gonna have a different kind of culture than if you go to a sexual health center.*

Adding improved search features would improve the app's ease of navigation and *convenience* for users. When discussing the mental health and substance use self-evaluation tools available on the app, participants wanted the app to link them to care when necessary:

> *I was gonna say it might be really helpful to have it directly linked to you where you can go through the diagnostic process. So, if you were tested with the depression or whatever, and you're like, "Okay, well, what can I do to get better? How can I help this?" If it had the tips, the infographic, self-help tips, you can do that, or you can be like, here's a psychiatrist you can go to if you want access to medication if it's really not going well for you. Because I really had no idea where to start with that when I got diagnosed with BPD and stuff. I had no clue what to do.*

To better help them use the provider locating feature, participants were interested in a rating and review system in which they could report their experiences with providers. This was perceived as a potentially effective method to avoid stigmatizing health care experiences.

> *Maybe after you set an appointment, there's a questionnaire and you say your experiences with that certain facility. I feel like if people within the app that have been there and use it, they know exactly what they were getting into.*

Another recommendation that came up across different focus groups pertained to internet *access* in rural areas. Dedicated offline functionality would make the app more usable and *convenient* for individuals with poor or fluctuating internet connections:

> *This might sound a little bit weird, but a very good/dedicated offline feature, at the very least for the FAQS and stuff like that, and maybe even the quizzes. 'Cause a lot of rural areas don't have the best access to internet. Even with mobile data, sometimes it's difficult to get a good connection and things can fail.*

Discussion

Principal Findings

We sought to understand the facilitators of and barriers to the uptake of sexual health services among sexual and gender minority individuals living in the rural southern United States and their interest in and willingness to use a mobile app to access HIV and STI prevention information and telehealth services. In focus groups with cisgender men who have sex with men and transgender and gender-expansive populations in the rural southern United States, participants reported frequent barriers to receiving appropriate sexual health services and a high interest in telehealth and, specifically, mobile apps to access HIV or STI prevention information and services.

Participants described a variety of experiences with health care providers, with most participants reporting stigmatizing experiences. Although some participants were able to access clinics that specifically cater to the LGBTQ+ population, many had to travel substantial distances to these clinics. PrEP providers have been documented to be clustered around urban areas, with many people in rural areas living in PrEP deserts that require travel of 1 hour or more to reach a PrEP provider [36]. This lack of proximity to needed HIV prevention services highlights the need for alternative methods for accessing health care, including telehealth, which has been found to be acceptable to rural GBMSM [19,20].

Participants held overwhelmingly positive views of the Combine app and expressed a high degree of willingness to use it to access HIV or STI testing and other HIV prevention services. The app was viewed as an efficient and effective method for overcoming the several barriers to accessing sexual health care discussed earlier. Although some of the barriers to living in rural areas are shared, GBMSM and transgender and gender-expansive communities are not homogenous. Interventions that have been found to be acceptable to GBMSM in densely populated areas, such as the several HIV prevention apps currently being tested, might not be transportable to rural communities. Our results indicate that app-based interventions might be transportable, but adaptations might be necessary to make them acceptable. Although discretion is always a concern when planning interventions for marginalized communities, the concerns of rural sexual and gender minority individuals are heightened because of the increased insularity of the broader communities in which they live. Indeed, rural men who have sex with men have been found to have concerns about privacy and confidentiality in technology-based HIV prevention research studies; however, these concerns were not perceived as insurmountable barriers to participation [28]. In addition, participants indicated a lack of access to basic information about sex and sexual health care such that additional information and resources might be necessary to include.

Drawing on their experiences of living in rural areas, participants had several suggestions for content that should be included in the app. For example, participants discussed the lack of relevant sexual health education that LGBTQ+ students in rural areas receive, a barrier that has been noted elsewhere [37]. To compensate for this, participants suggested that an app for rural sexual and gender minority users should include basic information about sex, the risk of HIV and STIs, and options for reducing the risk of HIV and STI transmission. Participants highlighted that multimedia presentations of this information (eg, diagrams or animations of how to properly use an external condom) would maximize its utility. Inclusive language with respect to anatomy was also mentioned as a priority. This highlights the potential need to include responsive design elements that would allow a user to specify their preferred terminology for different parts of their anatomy (eg, front hole, vagina) upon first use so that language can be used appropriately in the app.

Participants also had several functional recommendations to improve the app, both in general and to increase its utility for rural users. Participants wanted the Combine app to be able to integrate with other apps on their phone. Currently, Combine integrates mapping apps to provide directions to testing locations identified within the app. However, Combine also includes a feature that allows users to set one-time or recurring reminders for different prevention activities (eg, HIV or STI testing and ordering condoms). Participants suggested that integrating this feature with calendar apps on their phone, which they use more frequently, would make this feature more valuable. Future adaptations of the app should incorporate as many app

integrations as feasible to increase the utility of the Combine app. The suggestion to make most of the app's functions available offline would also increase its utility for rural residents with poor cellular coverage.

Participants expressed multiple motivations and methods that contributed to their decision to trust a source of information. The lack of advertising on a provider's website was perceived to legitimize their services by demonstrating that they were not dependent on external support. Similarly, affiliations with trusted organizations, such as research institutions, were perceived to confer legitimacy. Implementers should consider the extent to which endorsements from or affiliations with existing organizations might improve buy-in from potential end users.

Participants also wanted to be able to search for content within the app. Referencing a real estate app, one participant indicated that they would be able to find the information they wanted more quickly if they could search or use different filters to narrow down the information presented to them. This type of feature could be particularly useful for experienced users who revisit the app to search for particular information.

A requested feature that participants suggested would be particularly useful for sexual and gender minority users in rural areas is a ratings and review system. To supplement the provider locator within the app, participants wanted the ability for users to rate providers and write reviews of their experiences. These ratings could be implemented in several ways. For example, ratings could be provided based on the overall experience or for certain aspects of the experience (eg, aspects of culturally competent care). Ratings could also potentially be presented based on the identity of the user providing the rating (eg, cisgender gay men and transgender women) so that other users could view the ratings most relevant to their own experience. Implementing this type of system, however, would require a substantial amount of effort to moderate, and rules would have to be generated for when to display user-submitted ratings and reviews; for example, after a certain number of ratings have been received for a given provider.

Limitations

These focus groups were a convenience sample of sexual and gender minority participants recruited via the web. Thus, they are not representative of all sexual and gender minority rural residents in the southern United States. Willingness to use a mobile HIV prevention app among those who are or are not online and do not volunteer to participate in research studies might differ from those who do. Participants also did not have the opportunity to interact with the Combine app; therefore, they were unable to provide detailed feedback on its specific functionality.

Conclusions

The results of these focus group discussions indicate that sexual and gender minority individuals in the rural southern United States could benefit from HIV prevention interventions delivered via mobile apps and that there is high interest in and willingness to use such apps. To meet the goal of the EHE initiative, all communities at risk of HIV must have access to HIV prevention and treatment services. Mobile apps might present an effective and scalable method for reaching sexual and gender minority individuals in rural areas, as they have already been shown to do in more densely populated locations [21,38,39].

Conflicts of Interest

None declared.

References

1. Priority jurisdictions: phase I. HIV.gov. URL: https://www.hiv.gov/federal-response/ending-the-hiv-epidemic/jurisdictions/phase-one [accessed 2022-02-01]
2. HIV surveillance in urban and nonurban areas 2018 (preliminary). Centers for Disease Control and Prevention. URL: https://www.cdc.gov/hiv/pdf/library/slidesets/cdc-hiv-surveillance-urban-nonurban-2018.pdf [accessed 2022-02-01]
3. de Voux A, Bernstein KT, Kirkcaldy RD, Zlotorzynska M, Sanchez T. Self-reported extragenital chlamydia and gonorrhea testing in the past 12 months among men who have sex with men in the United States-American men's internet survey, 2017. Sex Transm Dis 2019 Sep 11;46(9):563-570 [FREE Full text] [doi: 10.1097/OLQ.0000000000001032] [Medline: 31415039]
4. Sullivan PS, Sanchez TH, Zlotorzynska M, Chandler CJ, Sineath RC, Kahle E, et al. National trends in HIV pre-exposure prophylaxis awareness, willingness and use among United States men who have sex with men recruited online, 2013 through 2017. J Int AIDS Soc 2020 Mar;23(3):e25461 [FREE Full text] [doi: 10.1002/jia2.25461] [Medline: 32153119]
5. Rosenkrantz D, Black W, Abreu R, Aleshire M, Fallin-Bennett K. Health and health care of rural sexual and gender minorities: a systematic review. Stigma Health 2017 Aug;2(3):229-243. [doi: 10.1037/sah0000055]
6. Hubach RD, Currin JM, Sanders CA, Durham AR, Kavanaugh KE, Wheeler DL, et al. Barriers to access and adoption of pre-exposure prophylaxis for the prevention of HIV among men who have sex with men (MSM) in a relatively rural state. AIDS Educ Prev 2017 Aug;29(4):315-329. [doi: 10.1521/aeap.2017.29.4.315] [Medline: 28825858]
7. Owens C, Hubach RD, Williams D, Voorheis E, Lester J, Reece M, et al. Facilitators and barriers of pre-exposure prophylaxis (PrEP) uptake among rural men who have sex with men living in the midwestern U.S. Arch Sex Behav 2020 Aug 26;49(6):2179-2191 [FREE Full text] [doi: 10.1007/s10508-020-01654-6] [Medline: 32219687]
8. Sirota T. Attitudes among nurse educators toward homosexuality. J Nurs Educ 2013 Apr;52(4):219-227. [doi: 10.3928/01484834-20130320-01] [Medline: 23506172]

9. Petroll AE, Mosack KE. Physician awareness of sexual orientation and preventive health recommendations to men who have sex with men. Sex Transm Dis 2011 Jan;38(1):63-67 [FREE Full text] [doi: 10.1097/OLQ.0b013e3181ebd50f] [Medline: 20706178]
10. Stahlman S, Sanchez TH, Sullivan PS, Ketende S, Lyons C, Charurat ME, et al. The prevalence of sexual behavior stigma affecting gay men and other men who have sex with men across Sub-Saharan Africa and in the United States. JMIR Public Health Surveill 2016 Jul 26;2(2):e35 [FREE Full text] [doi: 10.2196/publichealth.5824] [Medline: 27460627]
11. Williams M, Bowen A, Horvath K. The social/sexual environment of gay men residing in a rural frontier state: implications for the development of HIV prevention programs. J Rural Health 2005;21(1):48-55 [FREE Full text] [doi: 10.1111/j.1748-0361.2005.tb00061.x] [Medline: 15667009]
12. Arscott J, Humphreys J, Merwin E, Relf M. "that guy is gay and black. That's a red flag." how HIV stigma and racism affect perception of risk among young black men who have sex with men. AIDS Behav 2019 Aug 08;24(1):173-184. [doi: 10.1007/S10461-019-02607-4]
13. Winter S, Diamond M, Green J, Karasic D, Reed T, Whittle S, et al. Transgender people: health at the margins of society. Lancet 2016 Jul;388(10042):390-400. [doi: 10.1016/s0140-6736(16)00683-8]
14. Safer JD, Coleman E, Feldman J, Garofalo R, Hembree W, Radix A, et al. Barriers to healthcare for transgender individuals. Curr Opin Endocrinol Diabetes Obes 2016 Apr;23(2):168-171 [FREE Full text] [doi: 10.1097/MED.0000000000000227] [Medline: 26910276]
15. Fennie KP, Trepka MJ, Maddox LM, Lutfi K, Lieb S. Comparison of individual and area level factors between HIV-infected cisgender and transgender individuals in Florida (2006-2014). AIDS Behav 2016 Oct 17;20(10):2186-2191 [FREE Full text] [doi: 10.1007/s10461-016-1308-y] [Medline: 26885811]
16. Seelman KL, Miller JF, Fawcett ZE, Cline L. Do transgender men have equal access to health care and engagement in preventive health behaviors compared to cisgender adults? Soc Work Health Care 2018 Aug 30;57(7):502-525. [doi: 10.1080/00981389.2018.1462292] [Medline: 29708468]
17. Demographics of mobile device ownership and adoption in the United States. Pew Research Center. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2019-11-19]
18. Demographics of social media users and adopters in the United States. Pew Research Center. URL: https://www.pewresearch.org/internet/fact-sheet/social-media/ [accessed 2022-05-16]
19. Hubach RD, O'Neil AM, Stowe M, Hamrick J, Giano Z, Currin JM. Preferred methods of HIV and sexually transmissible infection screening delivery among a rural sample of men who have sex with men. AIDS Patient Care STDS 2020 Nov 01;34(11):470-476. [doi: 10.1089/apc.2020.0170] [Medline: 33147083]
20. Hubach RD, O'Neil AM, Hamrick J, Ernst C, Stowe M, Stone T, et al. Assessing the amenability of Rural MSM to using telemedicine for medical and mental healthcare. Annals LGBTQ Public Population Health 2021 Jun 03;2(2):125-134. [doi: 10.1891/lgbtq-2020-0056]
21. Sullivan PS, Driggers R, Stekler JD, Siegler A, Goldenberg T, McDougal SJ, et al. Usability and acceptability of a mobile comprehensive HIV prevention app for men who have sex with men: a pilot study. JMIR Mhealth Uhealth 2017 Mar 09;5(3):e26 [FREE Full text] [doi: 10.2196/mhealth.7199] [Medline: 28279949]
22. Biello KB, Marrow E, Mimiaga MJ, Sullivan P, Hightow-Weidman L, Mayer KH. A mobile-based app (MyChoices) to increase uptake of HIV testing and pre-exposure prophylaxis by young men who have sex with men: protocol for a pilot randomized controlled trial. JMIR Res Protoc 2019 Jan 07;8(1):e10694. [doi: 10.2196/10694]
23. Liu A, Coleman K, Bojan K, Serrano PA, Oyedele T, Garcia A, et al. Developing a mobile app (LYNX) to support linkage to HIV/sexually transmitted infection testing and pre-exposure prophylaxis for young men who have sex with men: protocol for a randomized controlled trial. JMIR Res Protoc 2019 Jan 25;8(1):e10659 [FREE Full text] [doi: 10.2196/10659] [Medline: 30681964]
24. Liu AY, Laborde ND, Coleman K, Vittinghoff E, Gonzalez R, Wilde G, et al. DOT Diary: developing a novel mobile app using artificial intelligence and an electronic sexual diary to measure and support PrEP adherence among young men who have sex with men. AIDS Behav 2021 Apr;25(4):1001-1012 [FREE Full text] [doi: 10.1007/s10461-020-03054-2] [Medline: 33044687]
25. Jones J, Dominguez K, Stephenson R, Stekler JD, Castel AD, Mena LA, et al. A theoretically based mobile app to increase pre-exposure prophylaxis uptake among men who have sex with men: protocol for a randomized controlled trial. JMIR Res Protoc 2020 Feb 21;9(2):e16231 [FREE Full text] [doi: 10.2196/16231] [Medline: 32130178]
26. Sullivan PS, Zahn RJ, Wiatrek S, Chandler CJ, Hirshfield S, Stephenson R, et al. HIV prevention via mobile messaging for men who have sex with men (M-Cubed): protocol for a randomized controlled trial. JMIR Res Protoc 2019 Nov 15;8(11):e16439. [doi: 10.2196/16439]
27. Sullivan PS, Stephenson R, Hirshfield S, Mehta CC, Zahn R, Bauermeister JA, et al. Behavioral efficacy of a sexual health mobile app for men who have sex with men: randomized controlled trial of mobile messaging for men. J Med Internet Res 2022 Feb 02;24(2):e34574 [FREE Full text] [doi: 10.2196/34574] [Medline: 35025755]
28. Hubach RD, O'Neil A, Stowe M, Giano Z, Curtis B, Fisher CB. Perceived confidentiality risks of mobile technology-based ecologic momentary assessment to assess high-risk behaviors among rural men who have sex with men. Arch Sex Behav 2021 May;50(4):1641-1650 [FREE Full text] [doi: 10.1007/s10508-019-01612-x] [Medline: 32078710]

29. HIV surveillance report. Centers for Disease Control and Prevention. URL: http://www.cdc.gov/hiv/library/reports/hiv-surveillance.html [accessed 2021-08-01]
30. HUD USPS ZIP CODE Crosswalk Files. Office of Policy Development and Research (PD&R). URL: https://www.huduser.gov/portal/datasets/usps_crosswalk.html [accessed 2021-06-15]
31. Ingram D, Franco S. 2013 NCHS urban-rural classification scheme for counties. Vital Health Stat 2 2014 Apr(166):1-73 [FREE Full text] [Medline: 24776070]
32. Rural-urban commuting area codes. USDA Economic Research Service. URL: https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes/ [accessed 2021-06-15]
33. The Index of Relative Rurality (IRR)?: US County Data for 2000 and 2010. PURR. URL: https://purr.purdue.edu/publications/2960/1 [accessed 2021-06-15]
34. Glaser BG. The constant comparative method of qualitative analysis. Social Problems 1965 Apr;12(4):436-445. [doi: 10.2307/798843]
35. MISTR - PrEP, prescribed online and delivered to your door Internet. MISTR. URL: http://heymistr.com [accessed 2022-05-16]
36. Siegler AJ, Bratcher A, Weiss KM. Geographic access to preexposure prophylaxis clinics among men who have sex with men in the United States. Am J Public Health 2019 Sep;109(9):1216-1223. [doi: 10.2105/ajph.2019.305172]
37. Currin JM, Hubach RD, Durham AR, Kavanaugh KE, Vineyard Z, Croff JM. How gay and bisexual men compensate for the lack of meaningful sex education in a socially conservative state. Sex Educ 2017 Jul 19;17(6):667-681. [doi: 10.1080/14681811.2017.1355298]
38. Schnall R, Kuhns L, Pearson C, Bruce J, Batey DS, Radix A, et al. Preliminary results from a pragmatic clinical trial of MyPEEPS mobile to improve HIV prevention behaviors in young men. Stud Health Technol Inform 2020 Jun 16;270:1365-1366 [FREE Full text] [doi: 10.3233/SHTI200444] [Medline: 32570661]
39. Bauermeister JA, Tingler RC, Demers M, Connochie D, Gillard G, Shaver J, et al. Acceptability and preliminary efficacy of an online HIV prevention intervention for single young men who have sex with men seeking partners online: the Mydex project. AIDS Behav 2019 Feb 14;23(11):3064-3077. [doi: 10.1007/S10461-019-02426-7]

Abbreviations

EHE: Ending the HIV Epidemic
GBMSM: gay and bisexual men who have sex with men
LGBTQ+: lesbian, gay, bisexual, transgender, and queer+
mHealth: mobile health
PrEP: pre-exposure prophylaxis
STI: sexually transmitted infection

Edited by A Mavragani; submitted 17.03.22; peer-reviewed by R Hubach, S John; comments to author 13.05.22; revised version received 16.05.22; accepted 17.05.22; published 13.06.22.

Please cite as:
Jones J, Edwards OW, Merrill L, Sullivan PS, Stephenson R
Interest in HIV Prevention Mobile Phone Apps: Focus Group Study With Sexual and Gender Minority Persons Living in the Rural Southern United States
JMIR Form Res 2022;6(6):e38075
URL: https://formative.jmir.org/2022/6/e38075
doi:10.2196/38075
PMID:35699980

Original Paper

Remote Consulting in Primary Health Care in Low- and Middle-Income Countries: Feasibility Study of an Online Training Program to Support Care Delivery During the COVID-19 Pandemic

Andrew Downie[1,2*], BMedSci, MD, MPH; Titus Mashanya[3*], BA; Beatrice Chipwaza[3], BVetMed, MSc, PhD; Frances Griffiths[2,4], BSc, MBBS, MRCGP, PhD; Bronwyn Harris[2], MA, PhD; Albino Kalolo[3], MD, MSc, PhD; Sylvester Ndegese[3], BSIT; Jackie Sturt[5], RGN, RMN, PhD; Nicole De Valliere[6], BSc, MRES; Senga Pemba[3], MME, PhD

[1]School of Public Health and Preventative Medicine, Monash University, Melbourne, Australia

[2]Warwick Medical School, University of Warwick, Coventry, United Kingdom

[3]Department of Public Health, Faculty of Medicine, St Francis University College of Health and Allied Sciences, Ifakara, United Republic of Tanzania

[4]Centre for Health Policy, University of the Witwatersrand, Johannesburg, South Africa

[5]Florence Nightingale Faculty of Nursing, Midwifery and Palliative Care, King's College London, London, United Kingdom

[6]Warwick Clinical Trials Unit, University of Warwick, Coventry, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Frances Griffiths, BSc, MBBS, MRCGP, PhD
Warwick Medical School
University of Warwick
Medical School Building
Coventry, CV4 7HL
United Kingdom
Phone: 44 024 7657 4880
Email: F.E.Griffiths@warwick.ac.uk

Abstract

Background: Despite acceleration of remote consulting throughout the COVID-19 pandemic, many health care professionals are practicing without training to offer teleconsultation to their patients. This is especially challenging in resource-poor countries, where the telephone has not previously been widely used for health care.

Objective: As the COVID-19 pandemic dawned, we designed a modular online training program for REmote Consulting in primary Health care (REaCH). To optimize upscaling of knowledge and skills, we employed a train-the-trainer approach, training health workers (tier 1) to cascade the training to others (tier 2) in their locality. We aimed to determine whether REaCH training was acceptable and feasible to health workers in rural Tanzania to support their health care delivery during the pandemic.

Methods: We developed and pretested the REaCH training program in July 2020 and created 8 key modules. The program was then taught remotely via Moodle and WhatsApp (Meta Platforms) to 12 tier 1 trainees and cascaded to 63 tier 2 trainees working in Tanzania's rural Ulanga District (August-September 2020). We evaluated the program using a survey (informed by Kirkpatrick's model of evaluation) to capture trainee satisfaction with REaCH, the knowledge gained, and perceived behavior change; qualitative interviews to explore training experiences and views of remote consulting; and documentary analysis of emails, WhatsApp texts, and training reports generated through the program. Quantitative data were analyzed using descriptive statistics. Qualitative data were analyzed thematically. Findings were triangulated and integrated during interpretation.

Results: Of the 12 tier 1 trainees enrolled in the program, all completed the training; however, 2 (17%) encountered internet difficulties and failed to complete the evaluation. In addition, 1 (8%) opted out of the cascading process. Of the 63 tier 2 trainees, 61 (97%) completed the cascaded training. Of the 10 (83%) tier 1 trainees who completed the survey, 9 (90%) would recommend the program to others, reported receiving relevant skills and applying their learning to their daily work, demonstrating satisfaction, learning, and perceived behavior change. In qualitative interviews, tier 1 and 2 trainees identified several barriers to implementation of remote consulting, including lacking digital infrastructure, few resources, inflexible billing and record-keeping systems, and limited community awareness. The costs of data or airtime emerged as the greatest immediate barrier to supporting both the upscaling of REaCH training and subsequently the delivery of safe and trustworthy remote health care.

Conclusions: The REaCH training program is feasible, acceptable, and effective in changing trainees' behavior. However, government and organizational support is required to facilitate the expansion of the program and remote consulting in Tanzania and other low-resource settings.

(JMIR Form Res 2022;6(6):e32964) doi:10.2196/32964

KEYWORDS

remote consultation; mobile consulting; digital health; telehealth; mHealth; eHealth; mobile health; health care; cascade; train the trainer; low- and middle-income countries; rural areas; Tanzania; Kirkpatrick; consultation; training; low- and middle-income; rural; COVID-19

Introduction

Background

Essential health services are not available for over a third of the world's population, and most of this population is in low- and middle-income countries (LMICs) [1]. Marginalized communities, including those living in rural areas and informal settlements or slums, have least access to high-quality health care [2]. High-quality care includes appropriate and timely treatment and follow-up [2], and its provision forms part of the United Nations Sustainable Development Goal for health [3].

Even prior to the COVID-19 pandemic, remote consulting was considered to have the potential to increase access to quality health care, especially in rural communities [4-6]. It is estimated that 85% of individuals across LMICs own a mobile phone [7]. Although Tanzania has lower rates of ownership, it still has 75% mobile phone ownership across the population and 90% among health workers [8]. Mobile phone ownership is lower amongst rural, older, illiterate, and female populations compared to other population groups but is rapidly increasing [4,7]. Patients find remote consulting acceptable and appreciate the consistency and continuity of care achieved through improved communication [9].

From the beginning of the COVID-19 pandemic, the World Health Organization recommended remote consultation using phones or videoconferencing as an option for protecting the safety of patients and health workers and to enable continued health care provision [10,11]. Worldwide, in the face of the pandemic, remote consulting increased but often with little preparation and training [12]. This lack of training in the use of health technology is a key barrier to the acceptance and uptake of remote consulting in LMICs [4,13], along with health workers' worries about increasing personal workload [9].

Worldwide, continuing medical education delivered remotely has been shown to be acceptable, feasible, and desirable [14]. It enables greater geographic accessibility and time flexibility [15] and has been shown to be as effective as traditional teaching methods and far more effective than no training [16,17]. Issues of network connectivity, costs of data/airtime, access to electricity, and usability of the device are challenges that need to be addressed [18].

This paper first describes a remotely delivered education program called remote consulting in primary health care (REaCH) aimed at equipping health care workers in LMICs with knowledge, skills, and confidence to conduct remote consulting. We then present a 2-phase approach to evaluation: (1) a pretest phase to establish technical and face validity, and (2) our feasibility study of the delivery of the REaCH training to registered health workers and its cascade to other health workers, and the perceived impact of training on the delivery of health care remotely.

REaCH Training Program and Its Development

REaCH training aims to equip health workers with an understanding of the variety, benefits, challenges, and organizational changes associated with remote consulting and the skills for implementation of remote consulting in their health care facilities. The training was developed in partnership between St Francis University College of Health and Allied Sciences (SFUCHAS) (Tanzania), King's College London (KCL, UK), and the University of Warwick (UK). The REaCH training, and a sample presentation of the training materials, can be freely accessed on a not-for profit basis at the Warwick Medical School website [19].

The REaCH training, developed in April and May 2020, is designed for registered health workers (eg, nurses, doctors, clinical medical officers) with access to smartphones, at least intermittent access to Wi-Fi, and an ability to learn in English. We refer to these trainees as tier 1 trainees. They engage in self-directed learning using written and video materials. Activities and assignments are included, which encourage trainees to apply what they learn to their local context. Training materials are in English and can be downloaded as PDF files where network access is challenging. A facilitator introduces the 8-module course to the trainees and interacts with them via a social media platform to discuss the learning and assignments. Each module is designed to take 1-3 hours. The facilitator supports these tier 1 trainees to cascade their learning to health workers in their local team (tier 2 trainees) using the train-the-trainer approach. It is left to the discretion of the tier 1 trainees to decide what learning to cascade to the tier 2 trainees. Tier 2 trainees need to own a feature phone (ie, no internet or up to 2G enabled). In our pilot, the learning cascade was completed in the local language, Swahili.

The content of each module is described in Table 1. REaCH is delivered via Moodle [20], an open-source blended-learning app. For the facilitated discussions, in our pilot, we used WhatsApp (Meta Platforms) [21] as it was popular locally and content is encrypted; trainees did not share patient information on the group. An information and communication technology (ICT) officer provided telephone support to trainees when they encountered difficulty with Moodle and suggested solutions

when internet access was difficult (eg, travelling to a local village to download the materials).

We used the talent, resource, alignment, implementation, and nurturing (TRAIN) framework to optimize our train-the-trainer approach [22]. The facilitators who delivered the tier 1 training and the tier 1 trainees themselves were health professionals willing and able to train others (*talent*). We provided airtime and internet for facilitators, and each tier 1 trainee received £60 (US $74.30) for airtime and internet (*resource*). We provided tier 1 trainees with a certificate of course completion so they could add this to their training portfolio (*alignment*). Embedded within the REaCH training are teaching and activities related to implementation of remote consulting and how to cascade learning (*implementation*). There is opportunity for the tier 1 trainees to maintain contact on social media after the course for peer support (*nurturing*).

The facilitator is supported by a facilitator's guide incorporating pedagogical principles underpinning the course, logistics, expectations, and tips to optimize trainee engagement. The learner is provided with a guide covering learning expectations, how to seek help, how to organize cascade training, and other logistical issues.

In July 2020, we pretested the first iteration of the REaCH Moodle course to establish technical and face validity with university-based professionals, 11 from SFUCHAS and 1 from the United Kingdom. The test demonstrated that it was possible and acceptable to use Moodle for delivering the course.

Based on feedback from this test, we included the WhatsApp group for facilitator support, developed the facilitator and trainee guides, and notes on how to cascade each module, an introductory video, and the option of downloading course materials as PDF files to enable studying to continue when digital access was interrupted. This version of REaCH was used in the second iteration feasibility study (August 2020) described in this paper. During this period, we obtained funding to support the airtime requirements of leaners to undertake and cascade training and deliver remote consultations to their patients. This timeline is presented in Table 2.

Table 1. REaCH[a] modules.

Module	Description
Introduction	Why is remote consulting important?
1	What devices and platforms are used in remote consulting?
2	How does my role change and the care I provide my patients?
3	What are the risks and benefits of remote consulting?
4	What patient outcomes can I expect, including limiting COVID-19 spread?
5	What new issues arise in remote consulting that are different from face-to-face care?
6	What is my plan for delivering my work remotely (and that of my team/colleagues)?
7	How can I evaluate my own remotely delivered health care practice (and that of my team/colleagues)?
8	How can I influence others to change to remote consulting?

[a]REaCH: remote consulting in primary health care.

Table 2. Timeline: REaCH[a] development, training process, and feasibility study.

Period	Training phase	Details
April-May 2020	Training development	• REaCH training developed for online delivery (using Moodle and mobile devices)
July 2020	Pretesting	• REaCH training pretested with university-based professionals in Tanzania (n=11) and the United Kingdom (n=1) • Online training delivery mode found to be acceptable

July-August 2020	Adaptation	• REaCH training adapted to include:
		  • Use of WhatsApp
		  • Facilitator and trainee guides
		  • Cascade notes
		  • Introductory video
		  • Downloadable materials

August-September 2020[b]	Feasibility study (as reported in this paper)	• REaCH training delivered to 12 tier 1 trainees, Ulanga District • Training completed by 12 trainees • Evaluation completed by 10 (83%) trainees • Cascade training delivered by 9 (75%) trained tier 1 trainers to 63 tier 2 trainees
August 2020-March 2022	Trial	• Stepped-wedge trial of REaCH training in Tanzania and Nigeria underway

[a]REaCH: remote consulting in primary health care.

[b]Data collected and analyzed in this paper.

Feasibility Study Objectives

The feasibility study objectives were to evaluate the trainees':

- Response to REaCH training, their engagement levels, and their perceptions of the content and process (reaction)
- Perceptions of their level of understanding of the topic, including knowledge, skills, and attitudes, after undertaking the training (learning)
- Intended changes in how they deliver health care after completing the training and how the training contributed to this (behavior)

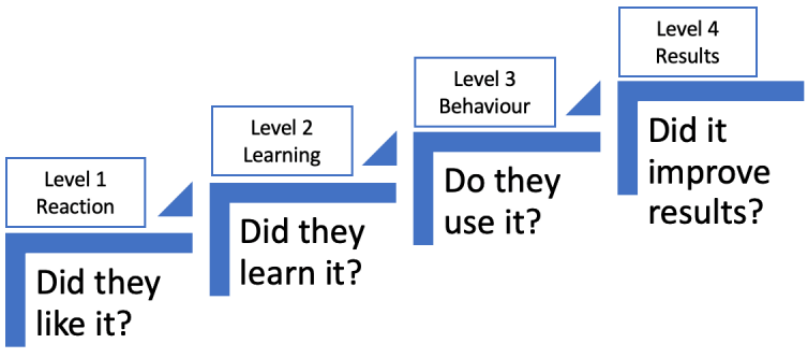
Methods

Study Design

In this feasibility study, we implemented and then evaluated the REaCH training using a survey, qualitative interviews, and documentary analysis. Our study was informed by Kirkpatrick's model [23] for assessing informal and formal learning (Figure 1).

We assessed the *reaction* and self-reported *learning* and intended *behavior* change.

Figure 1. Summary of 4 levels of Kirkpatrick's model.



Ethical Considerations

We used the Frascati definition of research, as summarized by the University of Warwick, to determine whether this study was considered research. We considered it not to be research, as its purpose was testing and standardization [24].

We subsequently checked our decision using the UK Medical Research Council and National Health Service (NHS) Health Research Authority tool for assessing ethics review, which indicated we did not need ethical review [25].

We received permission from the district medical officer (Ulanga District) for the participants (health care workers) to participate in the training and its evaluation.

Trainees and Setting

Tier 1 trainees were enrolled from health facilities in Ulanga District of remote rural Tanzania. Ulanga has a population of 265,203, with 1 hospital, 2 health centers, and 23 dispensaries [26,27]. Tier 1 trainees were selected using purposive and referral sampling and fulfilled the following criteria: they consulted with patients, worked in a rural area, owned a smartphone or computer, had access to Wi-Fi, were prepared to include remote consultations by phone as part of their health care practice, and were willing to cascade training to 7 other health workers in their team.

Tier 2 health workers were enrolled by the tier 1 trainees. They had to consult with patients, own a feature phone, and be prepared to add remote consultations by phone to their health care practice. The training was delivered between August 10 and September 2, 2020. All trainees received information about the evaluation and verbally consented to it.

Data Collection

The Survey

Questionnaires were developed by authors TM and SP for completion by trainees after each module and at the end of the training. These were structured around Kirkpatrick's model of learning [25], with Kirkpatrick's second and third levels (learning and behavior) assessed by self-report. We asked trainees about the process of undertaking the training (dichotomous questions and open-ended questions) and about their satisfaction (reaction), learning, and any intended changes to health care delivery as a result of the training (behavior) using 5-point Likert scales. A link to the survey was emailed to trainees and completed via SurveyMonkey [28].

Qualitative Interviews

The facilitator, all tier 1 trainees who completed the training, and, from each of their groups of tier 2 trainees, a convenience sample of 2 tier 2 trainees were invited for in-depth semistructured interviews. These were conducted by telephone following completion of the training by a researcher (TM) experienced in qualitative methods. Interviews explored participants' perceptions and experiences of the training and their views about remote consulting. Each interview was recorded digitally, transcribed verbatim, and translated by this researcher.

Documentary Analysis

We collated WhatsApp texts and emails between facilitator and trainees and reports written by tier 1 trainees after they had cascaded the training to tier 2 trainees. The tier 1 trainees reported on their experiences of cascading training, including topic selection, duration of training, preparedness for teaching and learning, how they motivated the tier 2 trainees, and advantages and disadvantages of the REaCH Moodle training approach.

Data Analysis and Trustworthiness

The survey results were analyzed descriptively. Interview transcripts were independently coded and analyzed thematically [29] by 3 team members (authors TM, BC, AD). Coding disagreements were resolved through discussion within the wider research team. TM analyzed the written documents thematically [29]. The research team held weekly debriefing meetings to reflect on the training and evaluation, identify/respond to challenges, share insights, and collectively make sense of the data [30]. We triangulated and integrated our findings in interpretation [31].

Results

Trainees

In total, 12 tier 1 trainees were enrolled within the REaCH training program, 3 (25%) women and 9 (75%) men. Tier 1 trainees were predominantly senior medical figures in participating health facilities (mostly doctors or assistant medical officers). In addition, 63 tier 2 trainees received cascaded training. The tier 2 trainees included a variety of health practitioners in the region: clinical officers, nurses, medical attendants, community health workers, and pharmacists, as well as 3 (5%) laboratory technicians and 2 (3%) radiologists, who were anecdotally delivering remote consulting. Of the 63 tier 2 trainees, 24 (38%) were women. Trainee characteristics are presented in Table 3. Of the 12 tier 1 trainees, all completed the training; however, 2 (17%) faced delays due to difficulty with internet connection and subsequently did not complete the evaluations or the cascading process. In addition, 1 (8%) tier 1 trainee faced personal circumstances, which precluded them from completing the cascading process. Thus, 9 (75%) tier 1 trainees cascaded their training to tier 2 health workers in their teams (N=63).

Table 3. Trainee characteristics.

Cadres	Tier 1 trainees			Tier 2 trainees	
	Enrolled (N=12), n (%)	Training completed (N=12), n (%)	Training cascaded (N=9), n (%)	Enrolled (N=63), n (%)	Training completed (N=61), n (%)
Medical doctors	6 (50)	6 (50)	5 (56)	1 (2)	1 (2)
Assistant medical officers	5 (42)	5 (42)	3 (33)	2 (3)	2 (3)
Clinical officers and assistant clinical officers	N/A[a]	N/A	N/A	18 (29)	18 (30)
Pharmacists	1 (8)	1 (8)	1 (11)	3 (5)	3 (5)
Community health workers	N/A	N/A	N/A	2 (3)	2 (3)
Radiologists	N/A	N/A	N/A	2 (3)	2 (3)
Laboratory technicians	N/A	N/A	N/A	3 (5)	3 (5)
Nurses	N/A	N/A	N/A	27 (43)	25 (41)
Medical attendants	N/A	N/A	N/A	5 (8)	5 (8)
Gender (female)	3 (25)	3 (25)	3 (33)	24 (38)	23 (38)

[a]N/A: not applicable.

The Survey: Training Process Questionnaires

Survey questions about the process of training were completed by 9 (75%) of 12 tier 1 trainees; 3 (25%) trainees were unable to complete the survey due to a poor internet connection.

All 9 responding tier 1 trainees had studied in their own personal time; 7 (78%) said that they also studied during working hours, 8 (89%) had completed all 8 REaCH modules, and 1 (11%) respondent completed 5 modules. All completed the assignments associated with the modules studied. Respondents spent 1-3 hours studying per module. Of the 9 respondents, 5 (56%) completed these modules in the allocated 6-day time frame, while the other 4 (44%) completed it within 8 days. Delays were due to busyness, device and network challenges, and initial low technological competence. All respondents found the assistance from the ICT officer and facilitators to be effective and timely.

The Survey: Reaction, Learning, and Behavior Questionnaires

Survey questions on reaction, learning, and behavior [23] were completed by 10 (83%) of 12 respondents (see Multimedia Appendix 1). There was little disagreement with the questions. All 10 respondents agreed that the training was useful and that facilitation was sufficient and timely. All 10 respondents appreciated the online and WhatsApp method of teaching and found that the learning outcomes were realistic and achievable. In addition, 9 (90%) of 10 respondents would strongly recommend this type of training to other health care workers.

Every respondent reported receiving the skills needed to learn remote consulting and to apply these skills to their jobs. They each reported already using the training in their daily work and being able to train other health care workers in this content.

Qualitative Interviews

Telephonic interviews were carried out with the tier 1 training facilitator, 9 (75%) tier 1 trainees, and 16 (25%) tier 2 trainees. Interviews lasted between 15 and 30 minutes each, with tier 1 interviews conducted in English and tier 2 interviews in Swahili. See Multimedia Appendix 2 for the interview question scaffold.

We present the results under the following themes: perceptions of the REaCH program, challenges encountered during the training, learning from REaCH training, how the training could be improved, and trainees' views on implementation of remote consulting into their routine practice. Trainees are labeled according to their tier of training and the order in which they were interviewed, as follows:

- Tier 1: participant A, B, C,…
- Tier 2: participant AA, BB, CC,….

Perceptions of the REaCH Training Program

Overall, the trainees appreciated the program and recommended continuation and expansion among their peers.

> *I wish to congratulate the initiators of this program. I would recommend this knowledge to be taught in the health colleges so that we now begin to recruit new doctors with high experience in remote consultation.* [Tier 1, participant D]

> *Generally, the participants perceived the course as a good course, something [that] is also a success.* [Tier 1, facilitator]

Challenges Encountered During the Training

Over half of the tier 1 trainees reported challenges with their digital technology, including storage capacity of smartphones, low technological competence, and network challenges.

> *There is no stable internet connection in this area, and this was 1 of the challenges I faced during the training. Ooh, likewise, the mobile network we use in our area is not stable.* [Tier 1, participant E]

Nevertheless, trainees found the assistance from ICT personnel and facilitators to be effective and timely.

> *There were several technical challenges and some issues concerning the arrangements of the modules; therefore, we used to seek instruction from the facilitator and ICT personnel, [and] actually, they were responding as soon as possible. We were told how to download the modules and the way we could go about reading them.* [Tier 1, participant A]

Although some trainees found it difficult to schedule the training around their work, others appreciated the flexibility of the online training.

> *The shortage of enough time…I spent many hours at work, so I had to make sure I read the modules in my extra time.* [Tier 1, participant D]
>
> *The training time planning was well arranged because it allowed us to engage in learning at any daily time.* [Tier 2, participant CC]

Learning From REaCH Training

All participating health care workers felt that their knowledge increased and that their behavior had changed since the training program. Some trainees were learning about remote consultation for the first time.

> *Yes, there are some changes; as you know, the modules have insisted on practicing remote consultation instead of face-to-face consultation, which we only trusted before. Recently, we have noticed that remote consultations are also appropriate, and actually, this alternative will work properly…!* [Tier 1, participant E]

Many remarked upon the usefulness of remote consulting during COVID-19.

> *Actually, this would assist much during this time of COVID-19 spread because it avoids the chance for having physical interaction between the patients and doctors.* [Tier 1, participant F]

Over half of the interviewees reported paying attention to privacy and confidentiality during remote consultations.

> *I do the remote consultation in a professional way by making sure I ask for consent, ensuring privacy as well as keeping their records, and making sure I continue to make follow-up on the patient's progress.* [Tier 1, participant A]

Trainees reported talking to patients before they attended clinic and following them up by phone rather than face-to-face. This included conducting remote consultations for patients who hesitated to attend face-to-face consultations out of fear of stigmatization.

> *I have started to offer advice to patients with shameful diseases remotely. You know the patients with gonorrhea can feel free to talk to a health care provider remotely rather than face-to-face.* [Tier 1, participant H]

Trainees had worked out how to bill for remote consultations.

> *I enjoyed this learning style because I have discovered that through this…this can give us an extra alternative to get money!* [Tier 1, participant D]

Recommendations for Improvement of the REaCH Training

Tier 1 trainees recommended a face-to-face meeting at the beginning of the course and additional time at the start to familiarize themselves with Moodle.

> *My advice is this…we should be making face-to-face meetings at least once at the beginning of a course that will be helpful in making things more clear! We can be taught physically on how to go through the Moodle and students' forum as well.* [Tier 1, participant D]

The facilitator agreed that additional time was needed at the start of the course for familiarization and supported the tier 1 trainees in including this during cascading. The facilitator supported some tier 1 trainees in producing printed materials for the tier 2 trainees. However, we found that many of the tier 2 trainees had smartphones and received the online materials easily.

> *The first thing that has been successful in this training [is] the majority of us received the learning materials on time simply because we have got smartphones through which we received them.* [Tier 2, participant MM]

The facilitator was keen to see the addition of incentives to engagement, such as accreditation and payment for time spent undertaking the training.

Implementation of Remote Consulting in Routine Practice

To further apply their learning in practice, trainees said they needed airtime and internet packages, suitable electronic devices, and improved infrastructure.

> *Our digital devices are not modern ones...we should be assisted with the internet packages to support online processes during the moment of interacting with the remote clients.* [Tier 1, participant B]
>
> *We can't provide the remote consultation if the supporting infrastructures like mobile networks are not working very well. Therefore, the government should ensure all necessary infrastructures for remote consultations.* [Tier 1, participant I]

Trainees emphasized the importance of governmental recognition to ensure adequate compensation for their work.

> *This needs some money…workers should be paid for this extra duty.* [Tier 1, participant D]

They recognized the need to inform the community about remote consulting.

> *Moreover, we need to make the community be aware and recognize this kind of consultation.* [Tier 1, participant I]

The lack of pharmacies and pathology laboratories in rural areas was identified as a barrier to successful remote consulting.

> *First of all, it will be difficult to make a physical examination, and the second challenge will be a shortage of pharmacies in remote areas, something [that] will make the remote clients fail to get medicines after consultation.* [Tier 2, participant CC]

> *The government should allow individuals to establish laboratories in remote areas. You know there are many laboratories that have been stopped due to the fact that they don't meet the eligibility requirements. So, we should have enough laboratories in rural areas so that clients may have test[s].* [Tier 2, participant DD]

Trainees noted that some members of the community would not have easy access to a phone, as they are owned by the heads of families and sharing phones can reduce confidentiality.

> *You know most of the mobile owners in the family level are heads of the families; therefore, the other family members will not be free enough to use those phones. So far, sharing phones will reduce the confidentiality of the clients' information.* [Tier 2, participant MM]

Older members of the community were unlikely to afford a phone, and there were community members who were illiterate and so unable to use text messaging.

> *Most of the community members, especially elders, are not possessing mobile phones, so they can't make consultations by themselves without asking the help from their neighbors.* [Tier 1, participant J]

> *Generally, this is a good idea, but I am doubting whether the elders will afford to pay the costs for remote consulting.* [Tier 2, participant NN]

> *So far, some of them are not aware of reading and writing; therefore, they can't send a text to the doctor when required to do so.* [Tier 1, participant J]

Some trainees were concerned with how to keep records of remote consultations.

> *We are still lacking the best alternative to keep the remote clients' records…we should find how to solve this challenge.* [Tier 1, participant B]

One said there needed to be a different way of referring patients between health care workers when they were consulting remotely.

> *To make referrals to remote clients, there should be an alternative for facilitation referrals from junior to senior HCWs [health care workers].* [Tier 1, participant L]

Qualitative findings from the interviews with tier 1 and 2 trainees were compared with analysis of the reports on the cascading process by tier 1 trainees.

Documentary Analysis: Tier 2 Training Process

Tier 2 training was completed over 3 days, with 2 pretraining preparation days where tier 1 trainees informed the tier 2 trainees about the aims of the course, its contents, and the training style and answered any questions. Tier 1 trainees selected modules 1, 2, 3, 5, and 7 for cascade as they were deemed to be the most clinically relevant (Table 1).

Introduction to training occurred via phone conferences and WhatsApp chats, and learning was primarily conducted through smartphone and featured phones with phone calls, texts, phone conferences, WhatsApp message group, and emails.

> *Soon after receiving the modules, the [tier 2] trainees started learning independently…when issues could not be understood, they used to make calls and send texts for more discussion and elaboration.* [Tier 1, participant E]

Where there were unstable internet connections, tier 2 trainees traveled to their nearest colleagues to pick up the module PDF files or to a nearby area with a stronger internet connection.

> *The participants from the areas with [an] unstable internet connection were advised to move to the areas with [an] internet connection in order to download the materials.* [Tier 1, participant A]

Of the 63 tier 2 trainees enrolled, 61 (97%) trainees completed the course, with 53 (87%) completing it within 3 days, while 2 (3%) did not complete the course due to personal reasons.

> *One participant's child got sick during the week of training that made her fail to complete the training in time.* [Tier 1, participant F]

Modifications used to ensure engagement included using reminder texts and phone calls to gauge and maintain attention of the trainees, using group discussions to increase teamwork, and conducting face-to-face conversations when the trainees and trainers were working in the same health facilities.

> *Sometimes, we were sending texts through the phone and WhatsApp media to remind them about the discussion time.* [Tier 1, participant D]

> *I also used to put some question in [the] WhatsApp platform to assess the trainees' understanding.* [Tier 1, participant F]

> *Participants' chats in WhatsApp assisted to assess the participation rate. You know, we were making calls to [the] training facilitator once per day to report on cascading progress and share the technical experience.* [Tier 1, participant B]

Modifications to solve logistical issues included translating the training documents into Swahili to overcome language barriers, providing downloadable materials that trainees could access from nearby villages when they had unstable internet connections, and moving group calls to early morning and evening hours to avoid working hours.

> *All in all, when I posted learning materials on the WhatsApp media, I tried to elaborate in Swahili to make them understand the contents.* [Tier 1, participant A]

Discussion

Principal Findings

This feasibility study found that remotely delivered professional REaCH training [19] using the Moodle app supported by cascade training infrastructures is technically and pedagogically feasible and well received by trainees in rural Tanzania. They were satisfied with the course and would recommend the program to other health care workers (reaction). They expressed that they learned skills needed to remotely consult within the health system, including how to bill patients for the consultations, and they were able to cascade the teaching (learning). Trainees reported confidently implementing remote consulting and increased understanding of topics, such as medical ethics of remote consulting and behavior change theory (behavior) [23].

Barriers to remote consulting implementation identified by our trainees include lacking digital infrastructure and few resources, inflexible billing and record-keeping systems, and limited community awareness about remote consulting. Having local technical support for learners proved invaluable to delivery and receipt of training. The greatest immediate barrier to supporting both the upscaling of REaCH training in LMICs and subsequently the delivery of safe and trustworthy remote health care is the cost of the data or airtime for the health workers themselves.

Comparison to Prior Work

Our REaCH training responds to a need identified by current research. In a systematic review of 14 studies assessing the feasibility and efficacy of remote consulting in LMICs, all studies identified that with adequate training, health care workers were able to learn to use mobile phones to deliver health care, but the review emphasized that sufficient initial and ongoing training is required to support the implementation of remote consulting [32]. In a systematic review of the barriers to remote consulting, lack of training was likewise identified as a key barrier [33]. Furthermore, during the COVID-19 pandemic, in a survey of physicians in Libya, 638 (94.8%) of 673 of participants expressed willingness to participate in a telemedicine training course [34].

Ediripulge et al. [35] literature review of 9 studies that described the delivery of training in telehealth not only emphasized the importance of adequate training to ensure integration of remote consulting in health systems but also found that the programs were predominantly conducted online and were a mixture of continuous professional development and university courses [35].

A scoping review, published after the development of the REaCH curriculum, describes the range of topics covered by courses that train health personnel for remote consulting [36]. Our course covered the key topics commonly taught, and included topics less commonly taught, including ethics, professionalism, and challenges of remote consulting. In this review, only 2 (5%) of 43 studies were conducted in LMICs [37,38]. One of these papers, similar to our study, evaluated its program using Kirkpatrick level 3 evaluation, while the other paper also included Kirkpatrick level 4 evaluation [39,40].

As in our study, the train-the-trainer approach to remote consulting education was successfully used in Rwanda to train community health care workers in monitoring pregnancy and pregnancy-related complications remotely and in Liberia to upskill traditional midwives to use mobile technology for short messaging service (SMS) texting [39,40]. Also replicated in our findings, remote delivery of remote consulting training in Brazil and India has been successful [38,41]. These trials and other similar remote consulting training programs in LMICs have been well received with high completion rates, as with our pilot study [42,43].

Strengths and Limitations

This study has some key strengths. We tested the program at several stages, undertook intensive evaluation at each stage, and were thus able to improve the program multiple times. We collected quantitative data and qualitative data to evaluate the training.

REaCH training and its pilot evaluation were undertaken at speed in response to the urgent need to support Africa's low-resource health care system in the face of the COVID-19 pandemic. Consequently, it has some limitations. The results are based on a relatively small number of health workers. Kirkpatrick's model informed the evaluation, but the second and third levels (learning and behavior) were assessed by self-report, with no external observation or validation. Although Kirkpatrick's model has its limitations for assessing medical education, it is useful for an evaluation, such as this that assesses immediate effects [44]. The survey questions were developed and delivered in a short time frame. Although each question captures 1 area of interest, some include 2 issues, which we are unable to tease out. There was a marked positive skew in the survey results, although in the exploratory semistructured interviews, respondents talked about both positives and negatives. The evaluation was conducted at 1 site, a single region in a single country, and by the team that developed the training.

We are currently running a stepped-wedge trial of REaCH training in Tanzania and Nigeria to evaluate actual *behavior* change and *results* in terms of the impact on health care delivery [45].

Conclusion

The REaCH program, providing training on remote consulting, is feasible and acceptable and successfully initiated behavior change in health care workers in a rural district in Tanzania. Trainees identified a need for resourcing of data/airtime and a technical and device infrastructure for the implementation of remote consulting.

Acknowledgments

We are deeply grateful to the health care workers who participated in the training and its evaluation. We thank the journal editors and 2 anonymous reviewers for the constructive, thoughtful comments on our manuscript. We also thank our colleagues in the Digital Innovation for Development in Africa (DIDA) project for the collegiality and shared insights into remote consulting.

This study was supported by funding from the UK Research and Innovation (UKRI) Global Challenges Research Fund for Digital Innovation for Development in Africa: Engineering and Physical Sciences Research Council (Grant EP/T030240/1) and the King's Together Fund: "How Can Remote Consulting Training for health workers in Rural Tanzania Be Optimised to Support Upscaling to Remote and Marginalised Communities of East and West Africa?"

Conflicts of Interest

The authors FG, BH, SP, and JS have an IP share in the licensing of REaCH training.

Multimedia Appendix 1
Survey responses aligned with Kirkpatrick's model (n=10 respondents).
[DOCX File , 18 KB - formative_v6i6e32964_app1.docx]

Multimedia Appendix 2
Interview questions scaffold.
[DOCX File , 16 KB - formative_v6i6e32964_app2.docx]

References

1. World Health Organization. Primary Health Care on the Road to Universal Health Coverage: 2019 Monitoring Report. Geneva: World Health Organization; 2019.
2. Kruk ME, Gage AD, Arsenault C, Jordan K, Leslie HH, Roder-DeWan S, et al. High-quality health systems in the Sustainable Development Goals era: time for a revolution. Lancet Global Health 2018 Nov;6(11):e1196-e1252 [FREE Full text] [doi: 10.1016/S2214-109X(18)30386-3] [Medline: 30196093]
3. United Nations. Transforming Our World: The 2030 Agenda for Sustainable Development. URL: https://sdgs.un.org/2030agenda [accessed 2022-05-11]
4. World Health Organization. Global Diffusion of eHealth - Making Universal Health Coverage Achievable: Report of the Third Global Survey on eHealth. Geneva: World Health Organization; 2017:154 pp.
5. Griffiths F, Watkins JA, Huxley C, Harris B, Cave J, Pemba S, et al. Mobile consulting (mConsulting) and its potential for providing access to quality healthcare for populations living in low-resource settings of low- and middle-income countries. Digit Health 2020;6:2055207620919594 [FREE Full text] [doi: 10.1177/2055207620919594] [Medline: 32341793]
6. Harris B, Ajisola M, Alam R, Watkins J, Arvanitis T, Bakibinga P. Mobile consulting (mConsulting) as an option for accessing healthcare services for communities in remote rural areas and urban slums in low- and middle- income countries: a mixed methods study. Digital Health 2021. [doi: 10.1177/20552076211033425]
7. Rowntree O. GSMA. Connected Women. The Mobile Gender Gap Report. London: GSMA; 2019.
8. Silver L, Johnson C. Majorities in Sub-Saharan Africa Own Mobile Phones, but Smartphone Adoption Is Modest. URL: https://www.pewresearch.org/global/2018/10/09/majorities-in-sub-saharan-africa-own-mobile-phones-but-smartphone-adoption-is-modest/ [accessed 2022-05-11]
9. World Health Organization. Recommendations on Digital Interventions for Health System Strengthening. URL: https://www.who.int/publications/i/item/9789241550505 [accessed 2022-05-11]
10. World Health Organization. Classification of Digital Health Interventions v1.0: A Shared Language to Describe the Uses of Digital Technology for Health. Geneva: World Health Organization; 2018.
11. World Health Organization. Operational considerations for case management of COVID-19 in health facility and community. Interim guidance. Pediatr Med Rodz 2020 Jul 14;16(1):27-32. [doi: 10.15557/pimr.2020.0004]
12. Webster P. Virtual health care in the era of COVID-19. Lancet 2020 Apr 11;395(10231):1180-1181 [FREE Full text] [doi: 10.1016/S0140-6736(20)30818-7] [Medline: 32278374]
13. Busagala L, Kawono G. Underlying challenges of e-health adoption in Tanzania. Int J Inform Commun Technol Res 2013;3(1):34-41.
14. World Health Organization. Web supplement 1: evidence-to-decision frameworks. WHO Guideline: Recommendations on Digital Interventions for Health System Strengthening. URL: https://apps.who.int/iris/bitstream/handle/10665/311979/WHO-RHR-19.6-eng.pdf?ua=1 [accessed 2022-05-11]
15. Ruggeri K, Farrington C, Brayne C. A global model for effective use and evaluation of e-learning in health. Telemed J E Health 2013 Apr;19(4):312-321 [FREE Full text] [doi: 10.1089/tmj.2012.0175] [Medline: 23472702]
16. Ruiz JG, Mintzer MJ, Leipzig RM. The impact of e-learning in medical education. Acad Med 2006 Mar;81(3):207-212. [doi: 10.1097/00001888-200603000-00002] [Medline: 16501260]

17. Cook DA, Levinson AJ, Garside S, Dupras DM, Erwin PJ, Montori VM. Internet-based learning in the health professions: a meta-analysis. JAMA 2008 Sep 10;300(10):1181-1196. [doi: 10.1001/jama.300.10.1181] [Medline: 18780847]
18. Lall P, Rees R, Law GCY, Dunleavy G, Cotič, Car J. Influences on the implementation of mobile learning for medical and nursing education: qualitative systematic review by the digital health education collaboration. J Med Internet Res 2019 Feb 28;21(2):e12895 [FREE Full text] [doi: 10.2196/12895] [Medline: 30816847]
19. Griffiths F. Remote Consulting for Healthcare (ReaCH): Training Health Workers to Continue Health Care Delivery Digitally in Low Resource Countries. URL: https://warwick.ac.uk/fac/sci/med/research/hscience/sssh/research/lyncs/COVID19 [accessed 2021-08-12]
20. Moodle. URL: https://moodle.org/ [accessed 2021-05-15]
21. WhatsApp. URL: https://www.whatsapp.com/ [accessed 2021-05-15]
22. Mormina M, Pinder S. A conceptual framework for training of trainers (ToT) interventions in global health. Global Health 2018 Oct 22;14(1):100 [FREE Full text] [doi: 10.1186/s12992-018-0420-3] [Medline: 30348183]
23. Kirkpatrick D, Kirkpatrick J. Evaluating Training Programs: The Four Levels. Oakland, CA: Berrett-Koehler; 2006.
24. Cochrane C. Frascati Definition of Research. URL: https://warwick.ac.uk/services/ris/systems/dataquality/frascati/ [accessed 2022-02-25]
25. NHS Health Research Authority. Do I Need NHS REC Review?. URL: http://www.hra-decisiontools.org.uk/ethics/ [accessed 2022-02-21]
26. The Planning Commission Dar es Salaam, Regional Commissioner's Office Morogoro. Morogoro Region Socio-Economic Profile. URL: http://www.tzonline.org/pdf/Morogoro.pdf [accessed 2022-05-11]
27. National Bureau of Statistics, Ministry of Finance, Office of Chief Government Statistician, President's Office, Finance, Economy and Development Planning. 2012 Population and Housing Census: Population Distribution by Administrative Areas. URL: http://tanzania.countrystat.org/fileadmin/user_upload/countrystat_fenix/congo/docs/Census%20General%20Report-2012PHC.pdf [accessed 2022-05-11]
28. SurveyMonkey. URL: https://www.surveymonkey.co.uk/ [accessed 2022-05-11]
29. Boyatzis R. Transforming Qualitative Information: Thematic Analysis and Code Development. Thousand Oaks, CA: SAGE Publications; 1998.
30. McMahon SA, Winch PJ. Systematic debriefing after qualitative encounters: an essential analysis step in applied qualitative research. BMJ Global Health 2018;3(5):e000837 [FREE Full text] [doi: 10.1136/bmjgh-2018-000837] [Medline: 30233833]
31. Noble H, Heale R. Triangulation in research, with examples. Evid Based Nurs 2019 Jul 14;22(3):67-68. [doi: 10.1136/ebnurs-2019-103145] [Medline: 31201209]
32. Agarwal S, Perry HB, Long L, Labrique AB. Evidence on feasibility and effective use of mHealth strategies by frontline health workers in developing countries: systematic review. Trop Med Int Health 2015 Aug;20(8):1003-1014 [FREE Full text] [doi: 10.1111/tmi.12525] [Medline: 25881735]
33. Kruse C, Betancourt J, Ortiz S, Valdes Luna SM, Bamrah IK, Segovia N. Barriers to the use of mobile health in improving health outcomes in developing countries: systematic review. J Med Internet Res 2019 Oct 09;21(10):e13263 [FREE Full text] [doi: 10.2196/13263] [Medline: 31593543]
34. Elhadi M, Elhadi A, Bouhuwaish A, Bin Alshiteewi F, Elmabrouk A, Alsuyihili A, et al. Telemedicine awareness, knowledge, attitude, and skills of health care workers in a low-resource country during the COVID-19 pandemic: cross-sectional study. J Med Internet Res 2021 Feb 25;23(2):e20812 [FREE Full text] [doi: 10.2196/20812] [Medline: 33600350]
35. Edirippulige S, Armfield NR. Education and training to support the use of clinical telehealth: a review of the literature. J Telemed Telecare 2017 Feb;23(2):273-282. [doi: 10.1177/1357633X16632968] [Medline: 26892005]
36. Stovel RG, Gabarin N, Cavalcanti RB, Abrams H. Curricular needs for training telemedicine physicians: a scoping review. Med Teach 2020 Nov;42(11):1234-1242. [doi: 10.1080/0142159X.2020.1799959] [Medline: 32757675]
37. Zafar A, Belard JL, Gilani S, Murad F, Khan M, Merrell RC. The impact of curriculum on a national telehealth program. Telemed J E Health 2008 Mar;14(2):195-198. [doi: 10.1089/tmj.2007.0029] [Medline: 18361709]
38. Manjunatha N, Kumar CN, Math SB, Thirthalli J. Designing and implementing an innovative digitally driven primary care psychiatry program in India. Indian J Psychiatry 2018;60(2):236-244 [FREE Full text] [doi: 10.4103/psychiatry.IndianJPsychiatry_214_18] [Medline: 30166682]
39. Ngabo F, Nguimfack J, Nwaigwe F, Mugeni C, Muhoza D, Wilson DR, et al. Designing and implementing an innovative SMS-based alert system (RapidSMS-MCH) to monitor pregnancy and reduce maternal and child deaths in Rwanda. Pan Afr Med J 2012;13:31 [FREE Full text] [Medline: 23330022]
40. Lori JR, Munro ML, Boyd CJ, Andreatta P. Cell phones to collect pregnancy data from remote areas in Liberia. J Nurs Scholarsh 2012 Sep;44(3):294-301 [FREE Full text] [doi: 10.1111/j.1547-5069.2012.01451.x] [Medline: 22672157]
41. Carvalho VS, Picanço MR, Volschan A, Bezerra DC. Impact of simulation training on a telestroke network. Int J Stroke 2019 Jul;14(5):500-507. [doi: 10.1177/1747493018791030] [Medline: 30040047]
42. dos Santos A, Alves HJ, Nogueira JT, Torres RM, Melo MDCB. Telehealth distance education course in Latin America: analysis of an experience involving 15 countries. Telemed J E Health 2014 Aug;20(8):736-741. [doi: 10.1089/tmj.2013.0291] [Medline: 24901742]

43. Ishibashi Y, Juzoji H, Kitano T, Nakajima I. ITU e-health training program for Pacific Island community with the support of the Sasakawa peace foundation. J Med Syst 2011 Jun 12;35(3):345-351. [doi: 10.1007/s10916-009-9371-9] [Medline: 20703555]
44. Yardley S, Dornan T. Kirkpatrick's levels and education 'evidence'. Med Educ 2012 Jan;46(1):97-106. [doi: 10.1111/j.1365-2923.2011.04076.x] [Medline: 22150201]
45. Sturt J. Determining Trustworthiness and Safety of Remote Consulting in Primary Healthcare for Chronic Disease Populations in Africa. URL: https://www.kcl.ac.uk/research/reach-trial [accessed 2022-02-28]

Abbreviations

ICT: information and communication technology
LMIC: low- and middle-income country
REaCH: remote consulting in primary health care
SFUCHAS: St Francis University College of Health and Allied Sciences

Edited by T Leung; submitted 23.08.21; peer-reviewed by C Kruse, L Agrawal; comments to author 07.01.22; revised version received 04.03.22; accepted 25.03.22; published 14.06.22.

Please cite as:
Downie A, Mashanya T, Chipwaza B, Griffiths F, Harris B, Kalolo A, Ndegese S, Sturt J, De Valliere N, Pemba S
Remote Consulting in Primary Health Care in Low- and Middle-Income Countries: Feasibility Study of an Online Training Program to Support Care Delivery During the COVID-19 Pandemic
JMIR Form Res 2022;6(6):e32964
URL: https://formative.jmir.org/2022/6/e32964
doi:10.2196/32964
PMID:35507772

Original Paper

Fairness in Mobile Phone–Based Mental Health Assessment Algorithms: Exploratory Study

Jinkyung Park[1], MA; Ramanathan Arunachalam[2], MSc; Vincent Silenzio[3], MD; Vivek K Singh[1,4], PhD

[1]School of Communication & Information, Rutgers University, New Brunswick, NJ, United States

[2]Department of Computer Science, Rutgers University, New Brunswick, NJ, United States

[3]School of Public Health, Rutgers University, Newark, NJ, United States

[4]Institute for Data, Systems, and Society, Massachusetts Institute of Technology, Cambridge, MA, United States

Corresponding Author:
Vivek K Singh, PhD
School of Communication & Information
Rutgers University
4 Huntington Street
New Brunswick, NJ, 08901
United States
Phone: 1 848 932 7588
Email: v.singh@rutgers.edu

Abstract

Background: Approximately 1 in 5 American adults experience mental illness every year. Thus, mobile phone–based mental health prediction apps that use phone data and artificial intelligence techniques for mental health assessment have become increasingly important and are being rapidly developed. At the same time, multiple artificial intelligence–related technologies (eg, face recognition and search results) have recently been reported to be biased regarding age, gender, and race. This study moves this discussion to a new domain: phone-based mental health assessment algorithms. It is important to ensure that such algorithms do not contribute to gender disparities through biased predictions across gender groups.

Objective: This research aimed to analyze the susceptibility of multiple commonly used machine learning approaches for gender bias in mobile mental health assessment and explore the use of an algorithmic disparate impact remover (DIR) approach to reduce bias levels while maintaining high accuracy.

Methods: First, we performed preprocessing and model training using the data set (N=55) obtained from a previous study. Accuracy levels and differences in accuracy across genders were computed using 5 different machine learning models. We selected the random forest model, which yielded the highest accuracy, for a more detailed audit and computed multiple metrics that are commonly used for fairness in the machine learning literature. Finally, we applied the DIR approach to reduce bias in the mental health assessment algorithm.

Results: The highest observed accuracy for the mental health assessment was 78.57%. Although this accuracy level raises optimism, the audit based on gender revealed that the performance of the algorithm was statistically significantly different between the male and female groups (eg, difference in accuracy across genders was 15.85%; P<.001). Similar trends were obtained for other fairness metrics. This disparity in performance was found to reduce significantly after the application of the DIR approach by adapting the data used for modeling (eg, the difference in accuracy across genders was 1.66%, and the reduction is statistically significant with P<.001).

Conclusions: This study grounds the need for algorithmic auditing in phone-based mental health assessment algorithms and the use of gender as a protected attribute to study fairness in such settings. Such audits and remedial steps are the building blocks for the widespread adoption of fair and accurate mental health assessment algorithms in the future.

(JMIR Form Res 2022;6(6):e34366) doi:10.2196/34366

KEYWORDS

algorithmic bias; mental health; health equity; medical informatics; health information systems; gender bias; mobile phone

Introduction

Background

Various machine learning (ML) algorithms are increasingly being used to make crucial decisions previously made by humans. Whether they are involved in approving loans, granting college admissions, or identifying the need for additional health support, automated algorithms find patterns, predict outcomes, and make decisions that may have consequential impacts on individuals' lives [1]. Indeed, the dependency on algorithms has eased our lives by replacing subjective human decisions with ML algorithms. The movement toward the application of automated algorithms in the health domain was not an exception. For instance, the proactive assessment of an individual's mental health is essential for maintaining a healthy and well-functioning society [2]. Although this holds the promise of dramatically wider access to mental health care, it is also fraught with inequities that can often inadvertently be baked into the algorithmic prediction of mental health levels.

ML algorithms attempt to find the generalized pattern from the training data, and sometimes these algorithms can manifest inherent biases across demographic characteristics such as age, race, ethnicity, and gender. A reason for the existing biases can be explained by *negative legacy* [3] (ie, the absence of sufficient data for a particular demographic group). For example, giving loans mostly to higher-income groups in the past may result in disapproval of loans to lower-income groups by algorithms that were informed by historical data, resulting in potential damage to individuals belonging to lower-income groups.

Such biases can be especially deleterious if they are part of health care algorithms. For instance, a recent study by Allen et al [4] found that algorithms used to assess mortality scores exhibit differential accuracy across races, thereby increasing racial disparities in health care. Similarly, Gianfrancesco et al [5] demonstrated that algorithmic predictions based on electronic health records can discriminate against multiple demographic groups. In particular, Obermeyer et al [1] showed that existing algorithms do not adequately identify the need for health support for people of color.

Building on these trends, we move the discussion of algorithmic fairness to mobile mental health assessment algorithms, which have been increasingly used in recent times [6]. With >6 billion users, mobile phones are one of the most ubiquitous consumer devices in the world. Many of them (especially smartphones) have capabilities conducive to monitoring an individual's physical activity, location, and communication patterns, each of which has been connected to mental health in the past [7,8]. Thus, mobile phones hold significant promise as a platform for monitoring multiple indicators of mental health risks and improving long-term management and treatment delivery to people with mental health issues [7,9]. At the same time, the creation of phone data–based ML models without considering the aspects of justice and fairness could reify, amplify, and multiply existing health disparities for certain segments of society (eg, women). Considering the abovementioned factors, the main research questions (RQs) of this study were as follows:

- RQ1: Are mobile phone-based mental health algorithms susceptible to bias in terms of gender?
- RQ2: Is it possible to reduce the level of bias while maintaining high accuracy?

Related Work

Predicting Mental Health

Over the past few decades, mental health has typically been assessed based on self-reported surveys that involved sporadic sampling, most of which were initiated after some significant events had taken place in an individual's life. Recently, as the availability of mobile phone data has increased, several studies have suggested using mobile phone data to detect and predict mental health conditions. Wang et al [10] introduced a mobile phone sensing system to automatically infer mental well-being, including depression, stress, flourishing, and loneliness. The study reported that automatically sensed conversation, activity, mobility, and sleep were significantly associated with mental health outcomes. By collecting data from sensors in mobile phone users (eg, location, messages, and calls), a longitudinal study showed a relationship between users' routines and moods [11]. Another study also found that mobile phone–based features such as call count, call response rate, and the number of new contacts are positively associated with mental health [8]. Using location information collected by a mobile phone app, Canzian and Musolesi [12] assessed the correlation between mobility patterns and the presence of depressive mood. A similar study also presented the relationship between depressive symptoms and the use of mobile phones and the movement through geographic spaces [7].

The results of the abovementioned studies provide clear evidence of interconnections between mobile phone data features and mental health conditions. More importantly, they suggested the potential of developing phone-based ML algorithms as a basis for the unobtrusive prediction of mental health conditions. However, to the best of our knowledge, no study has examined the possibility of algorithmic bias in predicting mental health status by using mobile phone data. Motivated by previous work on algorithmic fairness (see the *Algorithmic Fairness* section), this study attempted to mitigate the discriminatory impact of gender on mental health prediction algorithms.

Algorithmic Fairness

An increasing amount of research has suggested that ML algorithms in many domains are not free from discriminatory decision-making. Even with the best intentions, data-driven algorithmic decision-making processes can reproduce, inherit, or reflect the existing social biases. Algorithmic bias may stem from different sources, including (1) input data that may have unequal representation from different groups, (2) an algorithm that has been inadvertently or knowingly coded to make unfair decisions, (3) misuse of certain models in a different context, and (4) biased training data, which reaffirms that social biases may be used as evidence that an algorithm performs well [13]. Broadly, the sociotechnical system framework underscores that the value system of the algorithm developers is coded during the algorithm design process; hence, each assumption (often

implicit) made by the developers influences the real-world performance of the algorithm [14].

At the same time, multiple bias mitigation techniques have been developed for fairness in the ML literature [15,16]. Roughly, they attempt to counter such algorithmic bias by modifying the training data (preprocessing), learning algorithms (in-processing), or prediction (postprocessing). Preprocessing approaches focus on adapting the data going into the algorithms [16], in-processing approaches change the core algorithm (eg, change optimization function) [15], and postprocessing algorithms tend to modify the predicted labels to increase fairness [17].

Despite the plethora of related work, attempts to ensure algorithmic fairness toward a protected attribute (gender in our case) in the algorithmic assessment of mental health (high or low) have not been made.

Gender Bias

Various attempts have been made to tackle the issue of gender bias in computer algorithms by auditing algorithms for gender bias and modifying algorithms to eliminate stereotypes. For example, a study found that image search results for occupations could amplify gender stereotypes by portraying the minority gender as less professional [18]. Another study found gender stereotypes in word embeddings (eg, a framework to represent text data as vectors) and created debiasing algorithms to reduce gender bias while preserving the utility of the embeddings [19]. Furthermore, Zhao et al [20] tackled the problem of the effect of data imbalance, arguing that such data imbalance can worsen discrimination in terms of gender. They quantified the biases in visual recognition models and calibrated the models to reduce bias. However, no research has been conducted on gender equality using classification algorithms that predict mental health.

This study addressed the problem of identifying and reducing gender bias, as the overrepresentation of men in training data could accelerate gender inequality in mental health prediction algorithms. Particularly, we focused on the issue of *negativelegacy*, as suggested by Kamishima et al [3], which involves the idea that unfair sampling or labeling in the training data may lead to a disparate impact [16,21] on a certain group of people (eg, granting loans mostly to those who with higher income in the past may result in disapproval of loans to those with low income by the algorithms).

Perspective on Fairness and Justice

There exist multiple interpretations of fairness in the algorithmic fairness literature [22]. For instance, scholars define fairness in terms of maximizing utility for groups or respecting various rules such as individual rights and freedoms [23,24]. However, other interpretations abound, some of which are mutually incompatible [25].

The most commonly used approaches are those based on *distributive* and *procedural* justice [22]. While distributive justice focuses on how outcomes are distributed across the population, procedural justice focuses on the processes used to undertake the decisions [26,27].

An influential philosophical theory of fairness is attributed to the 20th-century philosopher Rawls, who equated fairness and justice, arguing broadly that fairness is *a demand for impartiality* [21,22]. In this study, we followed the approach for distributive justice based on the interpretation of Rawls. Specifically, we considered an algorithm to be fair if its performance did not vary for individuals with different demographic descriptors (eg, gender).

This is related to the concept of *disparate impact* [28]. Disparate impact, in US labor law, refers to practices in areas such as employment and housing, which affect one group of people of a protected characteristic more adversely than another, even when the rules applied by employers or landlords appear to be neutral [29]. Most federal civil rights laws protect against disparate impacts based on race, color, religion, national origin, and sex as protected traits, and some laws include disability status and other traits.

Methods

Data Set

We used a labeled data set from a previous study by Singh and Long [8], which explored the associations between call log data and mental health based on a 10-week field and laboratory study. The data set included phone-based behavioral data and self-reported mental health survey data. Phone-based data (eg, call volume, interaction dynamics, diversity in contacts, tie strength, and temporal rhythms) were collected through the app installed on each participant's mobile phone. Meanwhile, mental health was measured via in-person survey sessions using the Mental Health Inventory subscale of the 36-Item Short Form Health Survey [30]. After passing a preprocessing and classification process, the study showed that automated ML algorithms using phone-based features achieved up to 80% accuracy in automatically classifying the mental health level (above or below the mean) of an individual [8].

A total of 59 participants completed the survey administered by Singh and Long [8]. However, some participants did not complete all the surveys, and some did not enter the correct identifier (International Mobile Equipment Identity [IMEI] number) consistently across surveys. This resulted in a subset of 45 participants in the study [8]. For this study, we returned to the survey data and decided to manually handle the *off-by-one* errors (ie, the mismatch in IMEI for different surveys only by 1 digit). Given that IMEI numbers have 14 to 15 digits, in the approximately 60-participant sample size, we considered the odds of 2 participants to be off by just 1 digit without human error being extremely low. This process helped us obtain a complete data set (ie, phone data, a mental health survey, and a demographic survey) for 55 participants.

The data set we obtained from Singh and Long [8] had gender as a demographic attribute that we considered a protected attribute. Note that a protected attribute in the algorithmic fairness literature is one on which performance should not depend [15]. Among these 55 participants, 21 (38%) self-reported their gender as women or female (minority class), and 34 (62%) self-described as men or male. Note that this study

does not differentiate between (biological) sex and (socially construed) gender. In addition, note that we consider the use of binary gender as a limitation of this study. Future studies should be conducted, which include participants with nonbinary gender identities.

Preprocessing and Model Training

The initial obtained data set was imbalanced (ie, there was not enough data for one class), which is a common problem in the fairness literature [31]. To mitigate the effect of imbalance, we applied the synthetic minority oversampling technique [32] to the training data (the test data remained in the original ratio). This technique works in balancing the data set by generating synthetic observations based on the existing minority observations.

Before moving on to the application of any ML algorithm, the missing values were filled with the median values of the corresponding features. To reduce the impact of features with high variance, the features were standardized by removing the mean and scaling to unit variance. To build a classification model for high or low mental health scores, instances were labeled into 2 categories (1=high and 0=low) via a median split.

With small sample data and high-dimensional space, there is always a chance of overfitting and reduced generalization. To avoid these issues, we used principal component analysis [33]. Principal component analysis confirmed that the top 5 components explained >99% of the variance (the larger the variation across a dimension, the more the information it contains); hence, we used the top 5 components as features for model creation.

The abovementioned latent features were passed to several classification algorithms to classify the level of mental health (ie, whether the score was above or below the mean score of the population). As the sample data size was relatively modest, we refrained from splitting the data set into training and test sets. Instead, as suggested by prior literature [8,34], we applied 5-fold cross-validations and experimented with 5 popular classification algorithms, including logistic regression, support vector machine, random forest, k-nearest neighbors, and multilayer perceptron neural networks using the *scikit-learn* library [35]. We ran all algorithms for 100 iterations, and the results are reported in the form of average overall accuracy, male accuracy (ie, accuracy for male individuals), and female accuracy (see the *Results* section).

Using the abovementioned data, we could, in principle, replicate the approach described by Singh and Long [8]. Although the features used were the same, we must note that the implementation was undertaken de novo with different preprocessing steps.

Auditing Mental Health Algorithms for Bias

Gender was selected as a protected attribute. Following the previous literature [36,37], men were considered the privileged group, and women were considered the unprivileged group. As there are multiple metrics to characterize accuracy in traditional ML (eg, observed accuracy, precision, recall, and F_1 score), past literature has discussed the need for multiple metrics to characterize bias in ML [13,31]. In this study, we adopted the five most commonly used metrics [15,16,38]:

1. Delta accuracy captures the difference in the accuracy of samples belonging to privileged and unprivileged groups based on sensitive features (eg, gender and race or ethnicity).
2. Delta true positive rate (ΔTPR) focuses on equal opportunity for truly deserving entries in both privileged and unprivileged groups to obtain a positive label (eg, higher mental health label) from the algorithm [13,15].
3. Delta false positive rate (ΔFPR) ensures that both the true positive rate and the false positive rate (instances where undeserving candidates are granted positive outcomes) are equal across different groups [15,39].
4. Statistical parity difference (SPD) calculates the difference in the probability of favorable outcomes from the algorithm being obtained by the unprivileged group to that of the privileged group [38].
5. Disparate impact captures the ratio of the probability of favorable outcomes for the unprivileged group to that of the privileged group [16] (see Multimedia Appendix 1 [13,15,16,39-41] for more details on the 5 metrics).

Following the principle of disparate impact, a fair information system is one in which the performance does not vary for individuals with different demographic descriptors (eg, gender); hence, the disparate impact metric should be close to 1.0. However, for practical settings, a model is considered biased if its value is <0.8 [40]. Meanwhile, the values of delta accuracy, ΔTPR, ΔFPR, and SPD should be close to zero in fair systems. Following the previous literature [39,41], we used a 2-tailed *t* test to assess whether there was a significant difference in accuracy, true positive rate, and false positive rate levels observed for the privileged and unprivileged groups.

Reducing Algorithmic Bias in Mental Health Assessment

Disparate impact remover (DIR) [16] is a preprocessing algorithm that modifies the feature values of the data set and makes the algorithm discrimination aware at the time of training. It does not require any changes in the classification algorithm, nor does it amend or postprocess the results of the classification algorithm, thus making it robust and applicable to different algorithms. The scenario in which DIR is needed to preprocess the data set depends on the metric called *balanced error rate (BER)*, defined as follows:

> BER = (error rate [S = privileged] – error rate [S = unprivileged]) / 2

In algorithmic fairness, the notion of BER is more important than the notion of traditional accuracy as, in most data sets, the contribution of the underprivileged attribute to the entire data set is lesser than that of the privileged attribute. For example, let us consider a data set with 100 rows, where 90 rows belong to the privileged group and 10 rows belong to the unprivileged group. With this data set, if the algorithm predicts all privileged rows right and unprivileged wrong, the error rate would be 10/100, which is 0.1, whereas the BER would be (0+1)/2, which is 0.5.

An approach discussed in the literature [16,17] is to replace the raw values of the data features with normalized variants that capture how extreme the value for an individual (eg, female) stands out within their own demographic group (eg, other women). In particular, the approach suggested by Feldman et al [16] tackles this issue by allowing the considered classes to have equal probabilities of scoring high values for any of the chosen features. With a toy example, where output is college admissions, input is Scholastic Assessment Test (SAT) scores, and with a binary notion of gender (men and women) for the protected class, this approach gives men and women separate scores based on their ranking within their own genders. For example, a man with an 80th percentile SAT score within the men's group is considered just as worthy as a woman with an 80th percentile SAT score within the women's group, irrespective of the actual SAT scores. In this way, the approach supports an equitable admission process across 2 genders. Note that in many practical settings, it is useful to undertake *partial repairs* (eg, move the scores at the same percentile across the privileged and unprivileged groups to be closer to each other rather than being congruent). Finally, the above approach can be extended to multidimensional input features for the algorithm. In the considered domain (phone-based mental health assessments), phone use patterns for men and women are known to differ [42,43]. Hence, using the same thresholds for the features (eg, number of phone calls) of men and women as symptoms of mental health issues could yield erroneous and biased results.

In this study, the DIR algorithm for bias reduction was implemented in Python using the IBM AIF360 library [15]. The algorithm was run 100 times, with each iteration having a shuffled version of the data set. The average results for the accuracy and fairness metrics are presented in the *Results* section.

Results

Mental Health Assessment Results

Table 1 shows the accuracy of multiple well-known ML algorithms for men and women (averaged over 100 iterations). The best-performing algorithm was random forest, which yielded 78.57% accuracy. These results are similar but not the same as those described by Singh and Long [8]. In both studies, the random forest algorithm yielded the best performance, and the highest observed accuracy was close to 80%. The random forest model with the highest accuracy had 100 estimators or number of trees in the forest and a maximum depth of 6.

Table 1. Results showing the average overall accuracy, accuracy for men, and accuracy for women for various machine learning models in mental health assessment (averaged over 100 iterations).

Machine learning models	Overall accuracy (%), mean (SD)	Male accuracy (%), mean (SD)	Female accuracy (%), mean (SD)	Delta across gender (%), mean (SD)	*P* value of the 2-tailed *t* test on delta
Multilayer perceptron neural networks	59.99 (3.67)	58.68 (8.14)	61.92 (9.24)	12.10 (10.41)	<.001
Support vector machine	63.17 (2.91)	65.98 (6.49)	59.60 (8.37)	12.20 (8.67)	<.001
Logistic regression	58.48 (2.69)	66.59 (5.47)	47.38 (6.75)	19.73 (9.80)	<.001
K-nearest neighbors	61.77 (1.78)	70.43 (3.72)	49.63 (5.89)	20.96 (8.46)	<.001
Random forest	78.57 (1.61)	87.16 (2.73)	71.31 (2.51)	15.85 (0.22)	<.001

Audit Results

We compared the accuracies of different algorithms for the male and female groups (Table 1). The performance was found to be significantly different for the 2 groups in each of the considered algorithms based on a nonpairwise (2-tailed) *t* test (α=.05; *P*<.001) [41]. This indicates that the commonly used ML algorithms, when used for phone-based mental health assessment, are susceptible to bias.

There, a trade-off is expected between accuracy and fairness (ie, with increased fairness, there is typically a dip in accuracy) [31], the random forest model with the highest observed accuracy was selected as the baseline model for further inspection of fairness.

For random forest, the average absolute delta accuracy was 15.85% (Table 2). The absolute values of ΔTPR and ΔFPR were 0.88% and 33.43%, respectively. The average SPD was 26.1%, and the average disparate impact was 0.682, which were distant from the ideal values of 0 and 1.0, respectively.

Table 2. The average score for bias metrics in the random forest–based mental health assessment algorithm (average of 100 iterations).

Bias metrics	Observed score, mean (SD)	Ideal score
Delta accuracy (%)	15.85 (0.22)	0
Delta true positive rate (%)	−0.88 (8.39)	0
Delta false positive rate (%)	33.43 (13.50)	0
Statistical parity difference (%)	26.1 (4.16)	0
Disparate impact	0.682 (0.049)	1.0

For 4 of the 5 considered metrics (ie, except ΔTPR), the fairness scores were far from the ideal scores. In other words, the developed model yielded significantly different outcomes for individuals across genders despite reasonable aggregate performance. More precisely, the model was mostly biased against the unprivileged group (in this case, women), and the disparate impact appeared to be a major issue.

Bias Reduction Results

We recomputed the abovementioned bias metrics after applying the bias reduction algorithm (DIR), and the results averaged over 100 iterations are reported in Table 3. Furthermore, a comparison of the results before and after applying the bias reduction algorithm is presented in Table 4.

Table 3. The average score for bias metrics after applying the disparate impact remover approach (average of 100 iterations).

Bias metrics	Observed score, mean (SD)	Ideal score
Delta accuracy (%)	1.66 (1.56)	0
Delta true positive rate (%)	3.74 (6.74)	0
Delta false positive rate (%)	5.58 (9.88)	0
Statistical parity difference (%)	−2.70 (1.71)	0
Disparate impact	1.09 (0.041)	1.0

Table 4. Comparison of delta accuracy, statistical parity difference, and disparate impact before and after applying the postprocessing algorithm.

Bias metrics	Baseline model, mean (SD)	After bias reduction, mean (SD)	Difference	*P* values of 2-tailed *t* test on delta
Delta accuracy (%)	15.85 (0.22)	1.66 (1.56)	14.19	<.001
Delta true positive rate (%)	−0.88 (8.39)	3.74 (6.74)	4.63	<.001
Delta false positive rate (%)	33.43 (13.50)	5.58 (9.88)	27.85	<.001
Statistical parity difference (%)	26.10 (4.16)	−2.70 (1.71)	28.80	<.001
Disparate impact	0.682 (0.049)	1.09 (0.041)	0.408	<.001

To test the significance of these improvements, we conducted a 2-tailed *t* test with α=.05 for each of the bias metrics for the before and after scores. The changes in all metrics were noteworthy (*P*<.001). The bias levels were reduced for 4 of the 5 metrics considered in this study. The only exception was ΔTPR, which was the only metric with a low (<5%) score in the baseline condition. This value remained <5% before and after the bias reduction process.

Note that as we move toward making the algorithm less biased, there is often a trade-off that arises in the form of the reduced overall accuracy of the model [13]. The accuracy levels for men and women were 87.16% and 71.31%, respectively (Δaccuracy 15.85%; mean 78.50%), before bias reduction. The accuracy levels changed to 78.49% and 76.83% for men and women, respectively (Δaccuracy 1.66%; mean 76.83%), after the bias reduction process. The 1.38% reduction (78.50%-77.12%) in the model accuracy was considered an acceptable loss in accuracy for the abovementioned improvements in fairness.

Discussion

Principal Findings

RQs of the Study

The first RQ in this work was as follows: are mobile phone–based mental health algorithms susceptible to bias in terms of gender?

As summarized in Table 1, we found statistically significant differences across genders in the performance of phone-based mental health assessment algorithms with an array of common ML algorithms. All of these point to the potential for disparate impact across gender with mental health assessment algorithms.

With respect to the performance of the highest accuracy algorithm (using random forest), we found noticeable differences in the performance of the algorithm across genders via the 5 commonly used bias metrics. As shown in Table 2, there was a difference in terms of all 5 metrics between the male and female groups. In particular, we found that the disparate impact ratio was 0.682 in the initial model. However, this value was much lower than the often recommended (and legally accepted) threshold of 0.8, irrespective of the intent of the designers [29]. Although the in-principle replications of algorithms described in the past literature may yield reasonable accuracy, their deployment will require them to meet the legal and ethical guidelines of disparate impact. In addition, similar fairness issues have been well studied in some other spaces (eg, policing and bank loans [44,45]); they are much less explored in algorithmic mental health assessment. However, they will become important with the increased deployment or adoption of mobile mental health tools.

The results also point to another domain in which women are disadvantaged. As per the US Department of Labor Statistics, women earn 82 cents for every dollar earned by men [46]. Similarly, recent research has reported worse performance for women in face recognition [47], Google Translate [48], and image search results [18]. The awareness of such disparities is an important first step in the creation of countermeasures. Broadly, such results in intersection with growing movements such as *Data Feminism* [49] can support the creation of more

equitable algorithms. Specifically, we hope that our findings will shed light on the need to ensure fairness in emerging mental health–related domains.

Finally, there are multiple potential reasons for the reduced performance of women in the considered algorithms. Given that the performance is consistently poorer for all the considered ML algorithms (Table 1), possible explanations may lie in the *negative legacy* and *data set imbalance*. Data imbalance is the lack of data samples from a particular demographic group for algorithms to learn from, and negative legacy refers to the lack of positive examples for algorithms to learn from for the unprivileged group [13,31]. For instance, Buolamwini and Gebru [47] argued that a lack of training samples is a reason for poorer performance for women and people of color. Similar to other areas, and perhaps even more urgently, there is a need for more diverse data samples to create accurate and fair ML models in mental health assessment algorithms.

The second RQ in this study was as follows: is it possible to reduce the level of bias while maintaining high accuracy?

On the basis of the results summarized in Table 4, we found that the DIR approach was effective in reducing the disparity in the performance of phone-based mental health assessment algorithms across genders. As reported in Table 4, there were statistically significant differences in terms of all 5 fairness metrics considered upon the application of the DIR approach.

Past literature has discussed the need for multiple metrics to characterize bias in ML [13,31] and that metrics can be orthogonal to each other [25,44]. A suggested process is for system designers to identify a set of parameters that they consider appropriate for a given task [50]. In this study, we considered disparate impact to be an important criterion, considered in consultation with the scores for other fairness metrics. In the considered scenario, noticeably large reductions in bias levels were observed regarding the 4 metrics, except for ΔTPR, where the scores were <5% before and after bias reduction. Finally, we noted that there was a 1.38% decrease in accuracy upon the application of the bias reduction approach.

Overall, we interpreted the results to imply that it is often possible to create fairer versions of algorithms. However, given the variety of fairness metrics that can be considered and the complexities of practical scenarios, the process of bias reduction is likely to involve a human-in-the-loop process and consideration of the trade-offs in terms of multiple metrics [50]. Hence, rather than identifying a silver bullet solution, there might be opportunities for multiple small modifications that allow fairer versions of the algorithms. Having said that, value-sensitive design needs to be an important part of the future design of similar applications [51], and algorithmic audits need to become an essential step in the process of medical approval of newer (algorithmic) diagnostic tools.

The obtained results have multiple implications for different stakeholders engaged in health information systems.

Health Informatics Researchers and Policy Designers

This study moves the conversation with health policy designers beyond the equity of the built environment (eg, access to hospitals and parks) to the equity of data infrastructure, which can profoundly influence the health outcomes for millions of individuals going forward [52]. Although there exist multiple legal and policy guidelines that counter the physical aspects of bias (eg, redlining [53]), there is relatively little work on legal and policy frameworks with digital algorithms that undertake similar roles.

Health Care Technology Companies

This study identified a feasible pathway for creating algorithms that balance accuracy and equity in the creation of novel health care applications. Hence, the findings support the creation of equitable versions of just-in-time mobile mental health intervention apps.

Health Care Providers

This study allows for more robust detection and flagging of mental health issues in patients. Fairer algorithms will reduce the odds of patients being flagged for interventions incorrectly simply because of demographic characteristics, thus allowing for the better alignment of resources between individual providers and the health care industry at large.

The Public

The ultimate goal of this study was to create and promote equity in mental health information technology. The fairness of algorithms is intimately connected with trust and adoption. In fact, recent research suggests that disparate impact diminishes consumer trust, even for advantaged users [40]. A robust fair detection process will allow for the scalable delivery of just-in-time and tailored mental health support services to a wider population. This is important, given the huge disparity between the need for mental health support and the percentage of the population that uses mental health services [54].

Limitations

This study has some limitations. It focused on a single data set with 55 individuals and considered a specific type of feature (phone data based, as described by Singh and Long [8] in the past literature). The use of binary gender in the assessment is another limitation of this study. Although this study examined many of the commonly used ML methods, other approaches are well represented in the literature. Hence, we will be cautious in generalizing the results until they are supported at a scale with samples of more representative populations and many other ML algorithms. Future work may also suggest other bias reduction techniques to reduce the discriminatory outcomes of mental health assessment algorithms based on protected attributes. At the same time, this work is the first empirical effort to analyze the difference in the performance of mental health assessment algorithms based on gender. A key contribution of this study is the motivation for future work in this domain using varied data sets and methods.

Conclusions

This study grounds the use of gender as a protected attribute to study fairness in phone-based mental health assessment algorithms. Mobile phones are now actively used by billions of individuals; hence, the automatic assessment of mental health using ML algorithms could potentially be beneficial in

estimating and intervening in billions of individuals' mental health conditions. An audit of commonly used ML algorithms for mental health assessment revealed that the performance of these algorithms can vary significantly depending on gender. This disparity in performance was found to be noticeably reduced after the application of a DIR approach by adapting the data used for modeling. The results move the literature forward on fairness in mental health assessment algorithms, particularly with gender as a protected attribute. Future work could consider larger data sets, protected attributes other than gender, and a newer approach to creating fair and accurate mental health assessment algorithms. Such results will pave the way for accurate and fair mental health support for all sections of society.

Acknowledgments

The authors would like to thank Gautam Sikka and Rahul Ellezhuthil for their help with the data analysis. This work was partly based on funding from the Rutgers Community Design for Health and Wellness Interdisciplinary Research Group, Rutgers Center for COVID-19 Response and Pandemic Preparedness, and Rutgers School of Communication & Information Scholarly Futures Grant.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The 5 metrics to measure bias in machine learning algorithms.
[DOCX File , 21 KB - formative_v6i6e34366_app1.docx]

References

1. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019 Oct 25;366(6464):447-453. [doi: 10.1126/science.aax2342] [Medline: 31649194]
2. Melcher J, Hays R, Torous J. Digital phenotyping for mental health of college students: a clinical review. Evid Based Ment Health 2020 Nov;23(4):161-166. [doi: 10.1136/ebmental-2020-300180] [Medline: 32998937]
3. Kamishima T, Akaho S, Asoh H, Sakuma J. Fairness-aware classifier with prejudice remover regularizer. In: Proceedings of the 2012 European Conference on Machine Learning and Knowledge Discovery in Databases. 2012 Presented at: ECML PKDD '12; September 24-28, 2012; Bristol, UK p. 35-50 URL: https://doi.org/10.1007/978-3-642-33486-3_3 [doi: 10.1007/978-3-642-33486-3_3]
4. Allen A, Mataraso S, Siefkas A, Burdick H, Braden G, Dellinger RP, et al. A racially unbiased, machine learning approach to prediction of mortality: algorithm development study. JMIR Public Health Surveill 2020 Oct 22;6(4):e22400 [FREE Full text] [doi: 10.2196/22400] [Medline: 33090117]
5. Gianfrancesco MA, Tamang S, Yazdany J, Schmajuk G. Potential biases in machine learning algorithms using electronic health record data. JAMA Intern Med 2018 Nov 01;178(11):1544-1547 [FREE Full text] [doi: 10.1001/jamainternmed.2018.3763] [Medline: 30128552]
6. Gindidis S, Stewart S, Roodenburg J. A systematic scoping review of adolescent mental health treatment using mobile apps. Adv Ment Health 2019;17(2):161-177 [FREE Full text] [doi: 10.1080/18387357.2018.1523680]
7. Saeb S, Zhang M, Karr CJ, Schueller SM, Corden ME, Kording KP, et al. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. J Med Internet Res 2015 Jul 15;17(7):e175 [FREE Full text] [doi: 10.2196/jmir.4273] [Medline: 26180009]
8. Singh VK, Long T. Automatic assessment of mental health using phone metadata. Proc Assoc Info Sci Technol 2018;55(1):450-459 [FREE Full text] [doi: 10.1002/pra2.2018.14505501049]
9. Abdullah S, Choudhury T. Sensing technologies for monitoring serious mental illnesses. IEEE MultiMedia 2018 Jan;25(1):61-75 [FREE Full text] [doi: 10.1109/mmul.2018.011921236]
10. Wang R, Chen F, Chen Z, Li T, Harari G, Tignor S, et al. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2014 Presented at: UbiComp '14; September 13-17, 2014; Seattle, WA, USA p. 3-14 URL: https://doi.org/10.1145/2632048.2632054 [doi: 10.1145/2632048.2632054]
11. Servia-Rodríguez S, Rachuri KK, Mascolo C, Rentfrow PJ, Lathia N, Sandstrom GM. Mobile sensing at the service of mental well-being: a large-scale longitudinal study. In: Proceedings of the 26th International Conference on World Wide Web. 2017 Presented at: WWW '17; April 3-7, 2017; Perth, Australia p. 103-112 URL: https://doi.org/10.1145/3038912.3052618 [doi: 10.1145/3038912.3052618]
12. Canzian L, Musolesi M. Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous

Computing. 2015 Presented at: UbiComp '15; September 7-11, 2015; Osaka, Japan p. 1293-1304 URL: https://doi.org/10.1145/2750858.2805845 [doi: 10.1145/2750858.2805845]
13. Lepri B, Oliver N, Letouzé E, Pentland A, Vinck P. Fair, transparent, and accountable algorithmic decision-making processes. Philos Technol 2018;31(4):611-627 [FREE Full text] [doi: 10.1007/s13347-017-0279-x]
14. Kuhlman C, Jackson L, Chunara R. No computation without representation: avoiding data and algorithm biases through diversity. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2020 Presented at: KDD '20; July 6-10, 2020; Virtual p. 3593 URL: https://doi.org/10.1145/3394486.3411074 [doi: 10.1145/3394486.3411074]
15. Bellamy RK, Dey K, Hind M, Hoffman SC, Houde S, Kannan K, et al. AI Fairness 360: an extensible toolkit for detecting and mitigating algorithmic bias. IBM J Res Dev 2019 Jul 1;63(4/5):4:1-415 [FREE Full text] [doi: 10.1147/JRD.2019.2942287]
16. Feldman M, Friedler SA, Moeller J, Scheidegger C, Venkatasubramanian S. Certifying and removing disparate impact. In: Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2015 Presented at: KDD '15; August 10-13, 2015; Sydney, Australia p. 259-268 URL: https://doi.org/10.1145/2783258.2783311 [doi: 10.1145/2783258.2783311]
17. Kamiran F, Karim A, Zhang X. Decision theory for discrimination-aware classification. In: Proceedings of the IEEE 12th International Conference on Data Mining. 2012 Presented at: ICDM '12; December 10-13, 2012; Brussels, Belgium p. 924-929 URL: https://doi.org/10.1109/icdm.2012.45 [doi: 10.1109/icdm.2012.45]
18. Singh VK, Chayko M, Inamdar R, Floegel D. Female librarians and male computer programmers? Gender bias in occupational images on digital media platforms. J Assoc Inf Sci Technol 2020 Jan 22;71(11):1281-1294 [FREE Full text] [doi: 10.1002/asi.24335]
19. Bolukbasi T, Chang KW, Zou J, Saligrama V, Kalai A. Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. In: Proceedings of the 30th International Conference on Neural Information Processing Systems. 2016 Presented at: NIPS'16; December 5-10, 2016; Barcelona, Spain p. 4356-4364. [doi: 10.5555/3157382.3157584]
20. Zhao J, Wang T, Yatskar M, Ordonez V, Chang KW. Men also like shopping: reducing gender bias amplification using corpus-level constraints. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. 2017 Presented at: EMNLP '17; September 7–11, 2017; Copenhagen, Denmark p. 2979-2989 URL: https://doi.org/10.18653/v1/d17-1323 [doi: 10.18653/v1/d17-1323]
21. Rawls J. A Theory of Justice: Revised edition. Cambridge, MA, USA: Harvard University Press; 1999.
22. Lee MK, Jain A, Cha HJ, Ojha S, Kusbit D. Procedural justice in algorithmic fairness: leveraging transparency and outcome control for fair algorithmic mediation. Proc ACM Hum Comput Interact 2019 Nov 07;3(CSCW):1-26 [FREE Full text] [doi: 10.1145/3359284]
23. Duster T. Individual fairness, group preferences, and the California strategy. Representations 1996;55:41-58 [FREE Full text] [doi: 10.2307/3043735]
24. Fish B, Bashardoust A, Boyd D, Friedler S, Scheidegger C, Venkatasubramanian S. Gaps in information access in social networks? In: Proceedings of the 2019 World Wide Web Conference. 2019 Presented at: WWW '19; May 13-17, 2019; San Francisco, CA, USA p. 480-490 URL: https://doi.org/10.1145/3308558.3313680 [doi: 10.1145/3308558.3313680]
25. Kleinberg J, Mullainathan S, Raghavan M. Inherent trade-offs in the fair determination of risk scores. In: Proceedings of the 8th Innovations in Theoretical Computer Science Conference. 2017 Presented at: ITCS '17; January 9-11, 2017; Berkeley, CA, USA URL: https://drops.dagstuhl.de/opus/volltexte/2017/8156/
26. Gummadi KP, Heidari H. Economic theories of distributive justice for fair machine learning. In: Proceedings of the 2019 World Wide Web Conference. 2019 Presented at: WWW '19; May 13-17, 2019; San Francisco, CA, USA p. 1301-1302 URL: https://doi.org/10.1145/3308560.3320101 [doi: 10.1145/3308560.3320101]
27. Ötting SK, Maier GW. The importance of procedural justice in Human–Machine Interactions: intelligent systems as new decision agents in organizations. Comput Human Behav 2018 Dec;89:27-39 [FREE Full text] [doi: 10.1016/j.chb.2018.07.022]
28. Barocas S, Selbst AD. Big data's disparate impact. Calif L Rev 2016;104:671 [FREE Full text] [doi: 10.2139/ssrn.2477899]
29. EEOC v. Sambo's of Georgia, Inc., 530 F. Supp. 86 (N.D. Ga. 1981): US District Court for the Northern District of Georgia. Justia US Law. 1981 Dec 30. URL: https://law.justia.com/cases/federal/district-courts/FSupp/530/86/1370384/ [accessed 2021-07-09]
30. Ware Jr JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992 Jun;30(6):473-483. [Medline: 1593914]
31. Pessach D, Shmueli E. A review on fairness in machine learning. ACM Comput Surv 2023 Apr 30;55(3):1-44 [FREE Full text] [doi: 10.1145/3494672]
32. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. J Artif Intell Res 2002 Jun 01;16:321-357 [FREE Full text] [doi: 10.1613/jair.953]
33. Monfreda M. Principal component analysis: a powerful interpretative tool at the service of analytical methodology. In: Sanguansat P, editor. Principal Component Analysis. London, UK: IntechOpen; 2012.

34. Dantas J. The importance of k-fold cross-validation for model prediction in machine learning. Towards Data Science. 2020 Nov 4. URL: https://towardsdatascience.com/the-importance-of-k-fold-cross-validation-for-model-prediction-in-machine-learning-4709d3fed2ef [accessed 2021-10-13]
35. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B. Scikit-learn: machine learning in Python. J Mach Learn Res 2011 Oct;12(2011):2825-2830 [FREE Full text]
36. Webster J. Shaping Women's Work: Gender, Employment and Information Technology. Milton Park, UK: Routledge; 2014.
37. Sarwono BK. Gender bias in a patriarchal society: a media analysis on virginity and reproductive health. Wacana 2012 Apr 01;14(1):37-60 [FREE Full text] [doi: 10.17510/wjhi.v14i1.48]
38. Dwork C, Hardt M, Pitassi T, Reingold O, Zemel R. Fairness through awareness. In: Proceedings of the 3rd Innovations in Theoretical Computer Science Conference. 2012 Presented at: ITCS '12; January 8-10, 2012; Cambridge, MA, USA p. 214-226 URL: https://doi.org/10.1145/2090236.2090255 [doi: 10.1145/2090236.2090255]
39. Alasadi J, Al Hilli A, Singh VK. Toward fairness in face matching algorithms. In: Proceedings of the 1st International Workshop on Fairness, Accountability, and Transparency in MultiMedia. 2019 Presented at: FAT/MM '19; October 25, 2019; Nice, France p. 19-25 URL: https://doi.org/10.1145/3347447.3356751 [doi: 10.1145/3347447.3356751]
40. Draws T, Szlávik Z, Timmermans B, Tintarev N, Varshney KR, Hind M. Disparate impact diminishes consumer trust even for advantaged users. In: Proceedings of the 16th International Conference on Persuasive Technology. 2021 Presented at: PERSUASIVE '21; April 12-14, 2021; Virtual p. 135-149 URL: https://doi.org/10.1007/978-3-030-79460-6_11 [doi: 10.1007/978-3-030-79460-6_11]
41. Hogg RV, McKean JW, Craig AT. Introduction to Mathematical Statistics. 6th edition. New York, NY, USA: Pearson; 2005.
42. Kimbrough AM, Guadagno RE, Muscanell NL, Dill J. Gender differences in mediated communication: women connect more than do men. Comput Human Behav 2013 May;29(3):896-900 [FREE Full text] [doi: 10.1016/j.chb.2012.12.005]
43. Forgays DK, Hyman I, Schreiber J. Texting everywhere for everything: gender and age differences in cell phone etiquette and use. Comput Human Behav 2014 Feb;31:314-321 [FREE Full text] [doi: 10.1016/j.chb.2013.10.053]
44. Chouldechova A. Fair prediction with disparate impact: a study of bias in recidivism prediction instruments. Big Data 2017 Jun;5(2):153-163. [doi: 10.1089/big.2016.0047] [Medline: 28632438]
45. Taylor WF. The ECOA and disparate impact theory: a historical perspective. J Law Policy 2018;26(2):575 [FREE Full text]
46. Jones J. 5 Facts About the State of the Gender Pay Gap. U.S. Department of Labor Blog. 2021 Mar 19. URL: https://blog.dol.gov/2021/03/19/5-facts-about-the-state-of-the-gender-pay-gap [accessed 2021-10-13]
47. Buolamwini J, Gebru T. Gender shades: intersectional accuracy disparities in commercial gender classification. In: Proceedings of 2018 Machine Learning Research Conference on Fairness, Accountability and Transparency. 2018 Presented at: PMLR '18; February 23-24, 2018; New York, NY, USA p. 1-15 URL: http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf
48. Prates MO, Avelar PH, Lamb LC. Assessing gender bias in machine translation: a case study with Google Translate. Neural Comput Applic 2020;32(10):6363-6381 [FREE Full text] [doi: 10.1007/s00521-019-04144-6]
49. D'Ignazio C, Klein LF. Data Feminism. Cambridge, MA, USA: MIT Press; 2020.
50. Noriega-Campero A, Bakker MA, Garcia-Bulle B, Pentland AS. Active fairness in algorithmic decision making. In: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society. 2019 Presented at: AIES '19; January 27-28, 2019; Honolulu, HI, USA p. 77-83 URL: https://doi.org/10.1145/3306618.3314277 [doi: 10.1145/3306618.3314277]
51. Friedman B. Value-sensitive design. interactions 1996 Dec;3(6):16-23 [FREE Full text] [doi: 10.1145/242485.242493]
52. Braveman P, Arkin E, Orleans T, Proctor D, Acker J, Plough A. What is health equity? Behav Sci Policy 2018;4(1):1-14 [FREE Full text] [doi: 10.1353/bsp.2018.0000]
53. Zenou Y, Boccard N. Racial discrimination and redlining in cities. J Urban Econ 2000 Sep;48(2):260-285 [FREE Full text] [doi: 10.1006/juec.1999.2166]
54. Augsberger A, Yeung A, Dougher M, Hahm HC. Factors influencing the underutilization of mental health services among Asian American women with a history of depression and suicide. BMC Health Serv Res 2015 Dec 08;15:542 [FREE Full text] [doi: 10.1186/s12913-015-1191-7] [Medline: 26645481]

Abbreviations

ΔFPR: delta false positive rate
ΔTPR: delta true positive rate
BER: balanced error rate
DIR: disparate impact remover
IMEI: International Mobile Equipment Identity
ML: machine learning
RQ: research question

SAT: Scholastic Assessment Test
SPD: statistical parity difference

Edited by A Mavragani; submitted 19.10.21; peer-reviewed by SM Ayyoubzadeh, E Sükei; comments to author 17.01.22; revised version received 27.03.22; accepted 10.04.22; published 14.06.22.

Please cite as:
Park J, Arunachalam R, Silenzio V, Singh VK
Fairness in Mobile Phone–Based Mental Health Assessment Algorithms: Exploratory Study
JMIR Form Res 2022;6(6):e34366
URL: https://formative.jmir.org/2022/6/e34366
doi:10.2196/34366
PMID:35699997

Original Paper

The Dispersion of Health Information–Seeking Behavior and Health Literacy in a State in the Southern United States: Cross-sectional Study

Nikita Rao[1*], MPH; Elizabeth L Tighe[2*], PhD; Iris Feinberg[3*], MA, MBA, PhD

[1]Mark Chaffin Center for Healthy Development, School of Public Health, Georgia State University, Atlanta, GA, United States

[2]Deparment of Psychology, Georgia State University, Atlanta, GA, United States

[3]Adult Literacy Research Center, Department of Learning Sciences, Georgia State University, Atlanta, GA, United States

[*]all authors contributed equally

Corresponding Author:
Nikita Rao, MPH
Mark Chaffin Center for Healthy Development
School of Public Health
Georgia State University
P O Box 3995
Atlanta, GA, 30302
United States
Phone: 1 404 413 1281
Fax: 1 404 413 1299
Email: nrao3@gsu.edu

Abstract

Background: The transmission of health information from in-person communication to web-based sources has changed over time. Patients can find, understand, and use their health information without meeting a health care provider and are able to participate more in their health care management. In recent years, the internet has emerged as the primary source of health information, although clinical providers remain the most credible source. The ease of access, anonymity, and busy schedules may be motivating factors to seek health information on the web. Social media has surfaced as a popular source of health information, as it can provide news in real time. The increase in the breadth and depth of health information available on the web has also led to a plethora of misinformation, and individuals are often unable to discern facts from fiction. Competencies in health literacy (HL) can help individuals better understand health information and enhance patient decision-making, as adequate HL is a precursor to positive health information–seeking behaviors (HISBs). Several factors such as age, sex, and socioeconomic status are known to moderate the association between HL and HISBs.

Objective: In this study, we aimed to examine the relationship between HL and HISBs in individuals living in a southern state in the United States by considering different demographic factors.

Methods: Participants aged ≥18 years were recruited using Qualtrics Research Services and stratified to match the statewide demographic characteristics of race and age. Demographics and source and frequency of health information were collected. The Health Literacy Questionnaire was used to collect self-reported HL experiences. SPSS (version 27; IBM Corp) was used for the analysis.

Results: A total of 520 participants met the criteria and completed the survey (mean age 36.3, SD 12.79 years). The internet was cited as the most used source of health information (mean 2.41, SD 0.93). Females are more likely to seek health information from physicians than males (r=0.121; P=.006). Older individuals are less likely to seek health information from the internet (r=−0.108; P=.02), social media (r=−0.225; P<.001), and friends (r=−0.090; P=.045) than younger individuals. Cluster analysis demonstrated that individuals with higher levels of HISBs were more likely to seek information from multiple sources than those with lower levels of HISBs (mean range 3.05-4.09, SD range 0.57-0.66; P<.001).

Conclusions: Age and sex are significantly associated with HISB. Older adults may benefit from web-based resources to monitor their health conditions. Higher levels of HL are significantly associated with greater HISB. Targeted strategies to improve HISB among individuals with lower levels of HL may improve their access, understanding, and use of health information.

(JMIR Form Res 2022;6(6):e34708) doi:10.2196/34708

KEYWORDS

health information–seeking behavior; health literacy; digital literacy; information retrieval; health literacy questionnaire

Introduction

Background

Health information–seeking behavior (HISB) is a complex construct that refers to the ways in which individuals seek information about health, illnesses, and health choices [1]. Armed with more knowledge, patients can participate more in their health care management, which has been shown to improve the effectiveness and efficiency of health care services and clinical outcomes [2,3]. Traditionally, health information is primarily communicated directly by health care providers to patients. Over time, information dissemination has transitioned from solely patient-provider interactions to obtaining information through Google searches, web-based communities, and one's social groups [4]. The predictors of HISB can be both contextual and personal, such as the environment, sociodemographic variables, or internal beliefs [5]. The advent of technologies including the internet, smartphones, and social media has prompted a drastic change in how people seek and access health information. Understanding how people seek health information in the digital age is a critical step in developing optimal health information delivery.

While digital health information provides a new landscape for HISB, clinical providers remain one of the most credible sources of health information, especially when it concerns major illnesses [6,7]. This is true irrespective of age or sex [6,7]. The relationship between the patient and provider is mediated by trust, which could be influenced by a patient's past experiences with their physicians and their perception of quality of care [8]. In addition, those with health insurance are more likely to communicate in-person with health care providers than those who are uninsured; one cause may be the inability to afford visits owing to high out-of-pocket expenses [9]. While clinical providers remain at the forefront of trustworthiness, the internet has emerged as the most frequently used source of health information [8]. Younger people, those with higher education, those with health insurance, being female, and people of color are more likely to use the internet to obtain health information than older people, those with lower education, the uninsured, being male, and White people [9,10]. Ease of access, anonymity, and busy schedules may be motivating factors to seek health information on the internet [10].

Health information from the internet also functions to supplement or cross-reference health information obtained elsewhere, such as from providers, especially in older adults as compared with younger adults [11]. Individuals also rely on the internet to stay abreast of new information in real time, with social media being one of the primary sources to stay informed in this manner. Social media use is most prevalent among millennials (those born between 1980 and 1995) and Gen Z (those born between 1996 and 2012) as a conduit of social support, communication, and obtaining and sharing all kinds of information, including health information [12,13]. The demographics of users on social media are similar to other web-based HISBs, in that being female, younger people, having higher levels of education, and people of color are using social media more than being male, older people, having lower levels of education, and White people [12,13]. In addition, social media's use of images, videos, and infographics provides information in a comprehensible and visually appealing manner [12]. Instagram, Snapchat, Facebook, and YouTube are popular platforms for seeking health information [12].

Health Information Seeking

While the younger generations are attuned to seeking health information on the web, the older generation maintains some habit of obtaining health information through traditionally printed materials such as books, newspapers, and magazines [13,14]. People who are older, have lower socioeconomic status, and have low digital literacy may rely on traditional media (printed materials, radio, or television) for health information [14]. Some studies show that non-English speakers were found to favor printed materials, which may be due to their availability in multiple languages [15]. This can pose a problem as health care providers are transitioning to web-based systems to disperse health information, rendering print sources as an antiquated form of health information dissemination [16]. Furthermore, research also shows that individuals with acute and episodic illnesses may also actively seek health information from friends and family [17]. However, when requiring information that is more specialized, such as with serious conditions such as cancer, stroke, and heart disease, friends and family are less frequently consulted [18]. Studies have shown mixed results with friends and family as both a frequent and infrequent source of health information [17,19]. Although friends and family may influence decision-making and provide social support, their role in HISB remains inconsistent.

The numerous methods of acquiring health information have also led to an increase in the breadth of information available. Both accurate and inaccurate information coexist, and there is no overarching regulation to ensure the validity or reliability of information. A study assessing the validity of the search terms "vaccine safety" and "vaccine danger" found that 55% of search results contained inaccurate information within the first 2 pages [19]. Another study on the accuracy of reproductive health information on the internet revealed that it took an average of 4 searches on a search engine to find relevant topics within the mass of media available on the web, indicating an inefficiency in finding health information (FHI) [20]. The wealth of social media platforms has only exacerbated this inefficiency. Although social media is a powerful medium for health information, individuals are now faced with information overload, uncertainty about the validity of their findings, misinformation, disinformation, and often conflicting information [21]. In this manner, information overload can cause stress and confusion for those experiencing it and can reduce their accuracy in health-related decision-making [21].

In the context of the COVID-19 pandemic, these stressors in HISB have been exacerbated by an onslaught of new and ever-evolving information from multiple scientific and nonscientific sources. The global scale and frequent evolution of COVID-19 has led to an infodemic or an overabundance of information [22]. The breadth and spread of misinformation within this context can sometimes lead to dangerous health consequences as misinformation is circulated and absorbed at a faster rate than accurate information [22], changing both people's understanding of COVID-19 as well as their potentially risky behaviors [23]. A lack of understanding of information related to COVID-19 has been associated with numerous negative health outcomes, including poor mental health, unwillingness to be vaccinated, and discontinuation of healthy preventive behaviors [24].

Although digital media often becomes the primary source of information during global health crises, the availability of many platforms and people's differential access and abilities in using these media are important factors in developing successful communication strategies to mitigate the risk of misinformation [25]. Digital health equity, including access to digital health information and language barriers, should also be considered when analyzing HISB [26]. Vulnerable populations that may not know how to access this information may require special attention and communication [25,26], as challenges associated with an individual's low health literacy (HL) and structural inequities can be exacerbated [27]. The proliferation and communication of health information during the pandemic has demonstrated the importance of the concept of HL by Nutbeam [28] as an important aspect in successfully combating information overload and correctly using health information [29].

Health Literacy and Health Information–Seeking Behaviors

The concept of HL has broadened over time, from a definition of understanding words and numbers in a medical context to the communication, understanding, and use of health knowledge in an interconnected manner [30]. Researchers agree that people should possess several competencies to find, understand, and use health information, including verbal, oral, decision-making, and numeracy skills [28]. Nutbeam [28] categorizes HL into 3 contexts: functional, interactive, and critical. Functional literacy is the ability to read and write, interactive literacy involves the application of health information to everyday circumstances, and critical literacy skills allow for the use of this information to exert control over life events [31]. Nutbeam [28] affirms the importance of distinguishing between these types of literacy and their practical applications. This ternary model of HL is used in several HL studies to map the development of HL competencies beyond the accumulation of basic health knowledge. This model has been used to develop competencies in a wide range of HL studies, including nutrition, patient decision-making, and children's HL practices [32-34]. Chin et al [35] promote a model of HL that encompasses processing capacity (working memory), general knowledge (ie, vocabulary), and specific health knowledge. This model, while structured differently than the model by Nutbeam [28], also emphasizes how these contexts are central to the overall accumulation of HL. Manganello [36] has expanded upon the model by Nutbeam [28] to include *media literacy*, which is defined as the ability to critically evaluate media messages. Zarcadoolas et al [37] also affirmed the importance of multilevel domains in understanding HL. While their model includes the constructs of *fundamental literacy, civic literacy, science literacy,* and *cultural literacy*, they posit that developing the public's HL needs to come from a multifaceted perspective [37].

With the rapid development of digital media, eHealth literacy has emerged as the use of information and communication technology to improve access to health care and health information [38]. eHealth literacy expands upon the traditional concept of HL and is associated with similar variables, such as age, education, income, culture, and experience using digital media [39]. Several studies and reviews have found that there is great potential for eHealth and eHealth tools as a manner of patient care and communication [38,40-42]. However, several challenges hinder its efficiency, including the type of technology used, the social environment, its evolving definition and measurements, and a lack of theoretical grounding in developing interventions [40]. In addition, the concept of eHealth literacy is significantly associated with an individual's level of overall HL, with higher overall HL positively correlated with greater eHealth literacy [38]. As such, while eHealth literacy is a topical and dynamic field, it is important to first evaluate the overall HL in a population as an antecedent to eHealth literacy.

The Health Literacy Questionnaire (HLQ) measures multidimensional, psychometric aspects of HL. By more than simply measuring whether an individual can read and write, the HLQ measures people's lived experiences of HL using self-reported experiences [43]. The HLQ comprises 9 scales that report patient-centered outcomes so that practitioners may be able to improve their interventions [43]. Studies in different samples from different populations have shown the HLQ to be an appropriate and strong measure of HL from a patient perspective [44-46], demonstrating strong internal consistency. Furthermore, the HLQ has been proven to allow for clinicians to better understand a patient's HL from their perspective and enable better communication and engagement between the patient and provider, leading to better health outcomes [46]. In addition, the HLQ scales can be organized into 3 categories in the Nutbeam model of HL [43]. Categorizing the HLQ scales into organizational schema in the Nutbeam model highlights how interventions for a specific domain under a scale can influence a broader concept of HL.

Of great importance in how people obtain and use information is how organizations provide health information. Organizational HL (OHL) plays an important role in information acquisition. OHL is the extent to which an organization's provision of health information is at a level where individuals can read, understand, and use it to make decisions about their health [47]. Successful OHL practices can lead to increased perceived quality and satisfaction of care for patients, ultimately leading to better clinical health outcomes [47,48]. Several guidelines provide evidence-based recommendations for OHL. Many interventions attempt to measure and develop individual HL, but it is of equal or greater significance that organizations include and measure HL in their overall planning for quality of care [49]. Lower

levels of HL may deter patient participation in care and can worsen the relationship between the patient and provider [48]. The ultimate purpose of this patient-centered approach is to support patients with their navigation of the health care system and access to, understanding of, and use of health information.

HL and HISB are inextricably linked, as the ability to seek, use, and comprehend health information requires a certain level of HL. In this way, adequate HL is a precursor to positive HISB [50]. Although significant correlations have been found between higher levels of HL and increased HISB, studies suggest that the link between HISB and HL is moderated by other factors, such as social networks, socioeconomic factors, and motivation to seek information [51]. Knowledge transfer within families and communities, social capital, and social engagement in the community contribute to an individual's level of both HL and HISB [40]. Age and sex are well-known factors that affect HL and HISB. Younger people and being female tend to have higher levels of HL, and their HISB differs significantly from those of older people and being male, respectively [13,16]. Being female and younger people are likely to find information more easily as compared with being male and older people [52]. HL has also been shown to decline with age and can lead to negative health outcomes, including increased mortality [53], although being female was associated with higher HL than being male in older age [40]. Individuals with higher levels of education are more likely to seek health information from a variety of sources [18] whereas those with lower levels of HL are less likely to do so [18]. In addition, those with lower levels of HL are more likely to need multiple sources of health information to be able to digest health knowledge and apply it [54].

Some studies show that individual characteristics, such as lower levels of income and lack of access to care associated with low HL are generally concentrated in rural areas, whereas those with higher levels of HL tend to reside in more urban areas [52]. However, these differences between urban and rural HL levels tend to disappear once factors such as age, sex, education, and income are accounted for [55]. As people in urban areas can also have low levels of HL, social and financial capital may be more strongly associated with HISB than HL itself [56]. Possible risk factors that may contribute to differences in urban and rural HL levels could be lack of employment, male sex, and language [57]. However, no single factor has been identified as significantly correlated with HL, and the differences between urban and rural levels of HL remain misunderstood. Significant associations have been found between urbanicity and HL and subsequent health outcomes, as those with higher levels of HL in urban areas are more likely to adopt healthy lifestyles [58]. Those characteristics of urbanicity associated with lower levels of HL are also correlated with HISB [59]. Urbanicity is also found to interact with HL to predict HISB. Chen et al [59] found that among rural residents, limited HL was associated with lower odds of access to health information, while among urban residents, HL was not associated with access to health information.

Study Context

Health disparities are exacerbated by macrosocial and microsocial factors, such as lack of health care access, low reading skills, health care costs, and geography. People living in the rural southern United States have higher rates of morbidity and mortality compared with their urban counterparts and those in other rural areas, with people of color experiencing higher rates of death and disease as compared with their White counterparts [60]. Most of these states have low rates of health care insurance, even with the expansion of Medicaid under the Affordable Care Act, leading to poorer health outcomes. In states that have chosen not to expand Medicaid, Texas, Georgia, and Florida account for half of the uninsured population [61]. Our study focused on a sample in Georgia, which ranks poorly for a variety of factors that influence HISB, including socioeconomic factors, health behaviors, and chronic disease [62]. Low-income and ethnic minorities are more likely to have poorer health outcomes such as heart conditions, cancer, stroke, and obesity than their White counterparts with higher income, exacerbating their negative health outcomes as they are unable to access health information in the traditional sense [62,63]. Understanding the structural and individual factors that influence health disparities in Georgia will allow us to understand the variables that influence HISB in this population as an essential first step in developing effective and accessible health information.

Goal of This Study

In this study, we examined the relationship between HL and HISB considering different demographic factors, such as age, sex, highest level of educational attainment, health insurance status, and county type as rural or urban. The HLQ was used to examine the lived HL experiences of individuals in a southern US state in terms of understanding, accessing, and using health information and health services as a measure of patient-reported outcomes. The purpose of this study was to examine HL and HISBs among a representative sample of adults in a southern US state by answering the following research questions:

- What are the average scores on the HISB scales?
- Are demographics (sex, age, education level, and county) related to the HISB scales?
- Are HISB scales predictive of HL outcomes (have sufficient information [HSI], critical appraisal [CA], FHI, and understanding health information [UHI])?
- Are there distinct clusters of participants based on HISB responses? Do these clusters differ by HL outcomes and participant demographics?

Methods

Recruitment

Participants who lived in Georgia and were aged ≥18 years were recruited using Qualtrics Research Services and stratified to match statewide demographic characteristics of geography and race (Explore Census Data [64]). Recruitment was conducted through email invitations or prompt survey platform prompts. Participants had individual incentive agreements with Qualtrics Research Services that included cash, gift cards, or retail store miles.

Measures

Demographic information on age, sex, race, highest level of educational attainment, health insurance status, and zip code was also collected. We assessed the frequency of sources of health information (from a lot to none) for printed materials, the internet, social media, physicians, and family and friends. We used the HLQ to collect different aspects of lived HL experiences related to accessing, understanding, and using health information: HSI, CA of health information, FHI, and UHI. All scales contain 4 to 6 items scored on a Likert-type scale; HSI and CA scales 1 to 4 have four response options (strongly disagree, disagree, agree, and strongly agree), and FHI and UHI scales 5 to 9 have five response options (cannot do, very difficult, quite difficult, easy, and very easy).

Statistical Analysis

We used SPSS (version 27; IBM Corp, 2020) for the analysis. Descriptive statistics included means, SDs, frequencies, and chi-square calculations. A 2-step cluster analysis was performed using the 4 HL scales.

Ethics Approval

The study was approved by the institutional review board of Georgia State University under the approval number H21522.

Results

Recruitment

Out of those who responded to the survey, 57.4% (520/905) met all the criteria and completed the survey. Table 1 presents the demographic characteristics of the participants. The mean age was 36.3 (SD 12.79; range 18-80) years. Racial categories matched state stratification rates. Educational attainment was split evenly between less than high school diploma, some college education, and college degree. Health insurance status showed that 72.6% (378/520) of the participants answered yes. Out of 520 respondents, 264 (50.8%) were urban county dwellers, which is similar to state demographics.

Table 1. Demographic data for overall sample (n=520).

Variable	Participants, n (%)
Female	371 (71.2)
Race	
Black or African American	167 (32.1)
White	301 (58)
Asian	28 (5.4)
Hispanic	13 (2.5)
Other	12 (2)
Educational attainment	
High school diploma or less	160 (30.7)
Some college education	177 (34)
College degree	184 (35.3)
Has health insurance	378 (72.6)
Geography	
Urban county	264 (50.8)
Rural county	256 (49.2)

Statistical Analysis

To address the first research question, “What are the average scores on the 5 health information–seeking behavior (HISB) scales?” means, SDs, and range values of the HISB scales are presented in Table 2. The scores ranged from 1 (none) to 4 (a lot).

To address the second research question, “Are demographics (sex, age, educational level, and county) related to the HISB scales?” Spearman rank correlations between demographics (sex, age, educational level, and county) and the 5 HISB scales are shown in Table 3. Of note, sex was significantly associated with the physician HISB scale (r_s=0.121; P=.01), such that being female was more likely to be associated with seeking health information from physicians. Age was significantly negatively associated with the internet (r=−0.108; P=.02), social media (r=−0.225; P<.001), and family and friends (r=−0.090; P=.045) HISB scales. This indicates that as age increases, individuals are less likely to seek health information from the internet, social media, and family and friends. There were no significant associations between educational level and county with any of the HISB scales (county with printed materials: r=−0.051, P=.25; county with IR: r=−0.062, P=.16; county with social media: r=0.010, P=.83; county with doctor or looks like physicians: r=−0.048, P=.28; county with family and friends: r=0.053, P=.23; education with printed materials r=−0.007, P=.88; education with IR: r=.033, P=.46; education with social media: r=0.057, P=.19; education with doctor or looks like physicians: r=0.045, P=.31; education with family and friends

r=0.023, *P*=.61). As expected, the 5 HISB scales were all positively and significantly related printed materials with IR: *r*=0.100, *P*=.02; printed materials with social media: *r*=0.331, *P*<.001; printed materials with doctor or looks like physicians: *r*=0.118, *P*=.007; printed materials with family and friends: *r*=0.221, *P*<.001; IR with social media: *r*=0.296, *P*<.001; IR with doctor or looks like physicians: r=0.203, *P*<.001; IR with family and friends: *r*=0.122, *P*=.005; social media with doctor or looks like physicians: *r*=0.095, *P*=.03; social media with family and friends: *r*=0.330, *P*<.001; doctor or looks like physicians with family and friends: *r*=0.274, *P*<.001).

To address the third research question, "Are HISB scales predictive of HL outcomes (HSI, CA, FHI, and UHI)?" a series of multiple regression analyses were conducted using SPSS. We ran 4 regression analyses, which included all HISB scales predicting each HL outcome (HSI, CA, FHI, and UHI; Tables 4-7). Before running the regression analyses, we dummy coded the 5 HISB scales. For each scale, 2 dummy codes were created and the "a lot" response served as the reference group: none or little (labeled D1) and some (labeled D2). We chose to dummy code the variables as the scale was ordinal and only ranged from 1 (none) to 4 (a lot). We considered these ordinal variables as categorical variables, which requires dummy coding. We collapsed the "none" and "little" groups into 1 group owing to low responses in these categories. As we were most interested in the comparison with the "a lot" group, we used it as the reference. These analyses helped us determine which HISB scales were most important for each HL outcome. For all 4 regression analyses, at least one dummy code for printed materials, internet, and physician HISB scales was uniquely predictive of all HL outcomes (see Tables 3-6 for estimates and unique R^2 values). Social media and family and friends HISB scales were not uniquely predictive of any of our HL outcomes. The HISB scales accounted for 0.223 to 0.331 of the variances in our 4 HL outcomes.

To address the fourth research question, "Are there distinct clusters of participants based on HISB responses? Do these clusters differ by HL outcomes and participant demographics?" we conducted a 2-step cluster analysis using SPSS with the 5 HISB scales (all scored 1-4). The results indicated that there were 2 distinct HISB clusters based on the 5 scales (Figure 1). We have descriptively labeled the clusters as "high HISB" (203/520, 39%) and "low HISB" (317/520, 60.9%). The pattern of responses to the HISB scales was similar across the 2 clusters; however, the high HISB cluster had more uniform and stable responses, with the exception of social media, which had the lowest reported health-seeking behavior (mean 1.95, SD 0.75). The low HISB cluster had printed materials followed by social media as their least frequent sources of health information (means 1.18 and 1.34, respectively, SD 0.38 and 0.58 respectively; Figure 1). We used chi-square difference tests to examine whether the clusters were differentiated based on 3 categorical demographics (sex, educational level, and county), and no significant differences were found (county, *P*=.87; education, *P*=.25; sex, *P*=.96). We used a 2-tailed *t* test to examine whether the clusters differed based on age, and there was no significant difference (*P*=.33). Finally, we conducted a series of 2-tailed *t* tests to examine whether the clusters differed based on our HL scales (HSI, CA, FHI, UHI, and actively managing health). There were significant mean differences for all 5 HL scales, such that the high HISB cluster had significantly higher means across all scales (means 3.05-4.09, SD 0.57-0.66; *P*<.001) than the low HISB cluster (means 2.63-3.78, SD 0.52-0.75; *P*<.001).

Table 2. Descriptives of health information–seeking behavior (HISB).

Variable	Value, mean (SD; range)
HISB printed materials	2.41 (0.93; 1-4)
HISB internet	3.28 (0.81; 1-4)
HISB social media	2.31 (1.00; 1-4)
HISB physicians	3.19 (0.82; 1-4)
HISB family and friends	2.78 (0.82; 1-4)

Table 3. Spearman rank correlations among the health information–seeking behavior (HISB) scales and demographics[a].

Variable	Sex	Age (years)	Education level	County	HISB print material	HISB internet	HISB social media	HISB physicians	HISB family and friends
Sex									
r	1	−0.038	0.018	0.081	−0.025	0.075	0.019	0.121	0.060
P value	__[b]	.40	.69	.07	.57	.09	.67	.006	.18
Age									
r	—	1	0.051	−0.028	−0.048	−0.108	−0.225	0.053	−0.090
P value	—	—	.25	.54	.29	.02	<.001	.24	.045
Education level									
r	—	—	1	−0.085	−0.007	0.033	0.057	0.045	0.023
P value	—	—	—	.05	.88	.46	.19	.31	.61
County									
r	—	—	—	1	−0.051	−0.062	0.010	−0.048	0.053
P value	—	—	—	—	.25	.16	.83	.28	.23
HISB print materials									
r	—	—	—	—	1	0.100	0.331	0.118	0.221
P value	—	—	—	—	—	.02	<.001	.007	<.001
HISB internet									
r	—	—	—	—	—	1	0.296	0.203	0.122
P value	—	—	—	—	—	—	<.001	<.001	.005
HISB social media									
r	—	—	—	—	—	—	1	0.095	0.330
P value	—	—	—	—	—	—	—	.03	<.001
HISB physicians									
r	—	—	—	—	—	—	—	1	0.274
P value	—	—	—	—	—	—	—	—	<.001
HISB family and friends									
r	—	—	—	—	—	—	—	—	1
P value	—	—	—	—	—	—	—	—	—

[a]The sample size ranges from 497 to 520. To interpret the direction of the correlations for dichotomous demographic variables, being female, some college or more, and rural county were all coded higher.

[b]Not applicable.

Table 4. Health information–seeking behavior scales predicting have sufficient information[a].

Predictor	Unique R^2	Coefficient (SE)	*t* test (*df*)	*P* value
Printed materials_D1[b]	0.056	−0.423 (0.070)	6.02 (508)	<.001
Printed materials_D2[c]	0.046	−0.381 (0.069)	5.50 (508)	<.001
Internet_D1	0.013	−0.130 (0.045)	2.89 (508)	.004
Internet_D2	N/A[d]	−0.078 (0.044)	1.77 (508)	.08
Social media_D1	N/A	−0.047 (0.068)	0.69 (508)	.49
Social media_D2	N/A	−0.029 (0.065)	0.44 (508)	.66
Doctor_D1	0.064	−0.294 (0.045)	6.47 (508)	<.001
Doctor_D2	0.011	−0.118 (0.044)	2.68 (508)	.008
Family and friends_D1	N/A	−0.056 (0.060)	0.93 (508)	.36
Family and friends_D2	N/A	−0.037 (0.058)	0.64 (508)	.52

[a]These are standardized coefficients. Total R^2=0.223. The response “a lot” served as the reference group for all dummy codes.

[b]D1: dummy code representing “none or little.”

[c]D2: dummy code representing “some.”

[d]N/A: not applicable.

Table 5. Health information–seeking behavior scales predicting critical appraisal[a].

Predictor	Unique R^2	Coefficient (SE)	*t* test (*df*)	*P* value
Printed material_D1[b]	0.072	−0.488 (0.067)	7.31 (508)	<.001
Printed material_D2[c]	0.037	−0.344 (0.066)	5.23 (508)	<.001
Internet_D1	0.043	−0.239 (0.042)	5.65 (508)	<.001
Internet_D2	0.019	−0.156 (0.042)	3.74 (508)	<.001
Social media_D1	N/A[d]	0.012 (0.064)	0.19 (508)	.85
Social media_D2	N/A	0.033 (0.061)	0.537 (508)	.59
Doctor_D1	0.070	−0.307 (0.043)	7.18 (508)	<.001
Doctor_D2	0.014	−0.135 (0.041)	3.25 (508)	<.001
Family and friends_D1	N/A	−0.057 (0.056)	1.00 (508)	.32
Family and friends_D2	N/A	−0.015 (0.055)	0.27 (508)	.79

[a]These are standardized coefficients. Total R^2=0.312. The response “a lot” served as the reference group for all dummy codes.

[b]D1: dummy code representing “none or little.”

[c]D2: dummy code representing “some.”

[d]N/A: not applicable.

Table 6. Health information–seeking behavior scales predicting finding health information[a].

Predictor	Unique R^2	Coefficient (SE)	*t* test (*df*)	*P* value
Printed material_D1[b]	0.018	−0.238 (0.068)	3.48 (509)	<.001
Printed material_D2[c]	N/A[d]	−0.125 (0.067)	1.85 (509)	.06
Internet_D1	0.022	−0.169 (0.044)	3.87 (509)	<.001
Internet_D2	N/A	−0.067 (0.043)	1.56 (509)	.12
Social media_D1	N/A	0.117 (0.066)	1.77 (509)	.08
Social media_D2	N/A	−0.106 (0.063)	1.68 (509)	.09
Doctor_D1	0.099	−0.365 (0.044)	8.26 (509)	<.001
Doctor_D2	0.057	−0.269 (0.043)	6.27 (509)	<.001
Family and friends_D1	N/A	−0.044 (0.058)	0.75 (509)	.45
Family and friends_D2	N/A	−0.038 (0.057)	0.68 (509)	.50

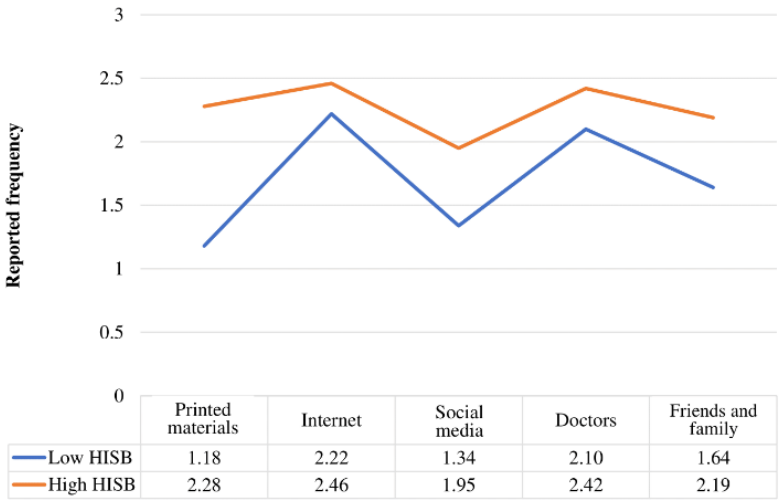
[a]These are standardized coefficients. Total R^2=0.263. The response “a lot” served as the reference group for all dummy codes.
[b]D1: dummy code representing “none or little.”
[c]D2: dummy code representing “some.”
[d]N/A: not applicable.

Table 7. Health information–seeking behavior scales predicting understanding health information[a].

Predictor	Unique R^2	Coefficient (SE)	*t* test (*df*)	*P* value
Printed material_D1[b]	0.020	−0.253 (0.069)	3.69 (509)	<.001
Printed material_D2[c]	0.016	−0.225 (0.068)	3.32 (509)	.001
Internet_D1	0.011	−0.121 (0.044)	2.76 (509)	.006
Internet_D2	N/A[d]	−0.031 (0.043)	0.71 (509)	.48
Social media_D1	N/A	−0.086 (0.066)	1.30 (509)	.19
Social media_D2	N/A	−0.086 (0.064)	1.36 (509)	.18
Doctor_D1	0.114	−0.391 (0.044)	8.81 (509)	<.001
Doctor_D2	0.073	−0.304 (0.043)	7.06 (509)	<.001
Family and friends_D1	N/A	−0.036 (0.059)	0.61 (509)	.54
Family and friends_D2	N/A	−0.029 (0.057)	0.51 (509)	.61

[a]These are standardized coefficients. Total R^2=0.256. The response “a lot” served as the reference group for all dummy codes.
[b]D1: dummy code representing “none or little.”
[c]D2: dummy code representing “some.”
[d]N/A: not applicable.

Figure 1. Two clusters based on 5 health information–seeking behaviors (HISBs).



	Printed materials	Internet	Social media	Doctors	Friends and family
Low HISB	1.18	2.22	1.34	2.10	1.64
High HISB	2.28	2.46	1.95	2.42	2.19

Discussion

Principal Findings

Our study highlights several important links between HISB and HL: (1) as age increases, people are less likely to seek health information from the internet and social media; (2) seeking health information from social media is not predictive of HL outcomes and is the least-used source of health information for people with high and low HL levels; and (3) people with high HL consistently exhibit more HISBs across multiple sources than those with low HL.

Although internet use has significantly increased in the past few years, disparities remain owing to age, gender, race, and socioeconomic status, which may persist in the digital gap between generations and among populations [65]. Studies show that people of all ages may prefer more traditional, printed medium or health care professionals for health information yet also seek corroboration, new information, or different perspectives from web-based HISBs [66]. Among those who use the internet, 79% have looked for health information of one kind or another, and 55% of these *online diagnosers* have spoken with a clinician about what they have found on the web [67]. Health information available on the web might be especially important for those with sensitive or stigmatized health issues such as drug use, unplanned pregnancies, and sexually transmitted diseases [68]. Older adults who may not have digital skills or digital access may not be able to access important health information, participate in decision-making with their health providers, reach provider websites to access patient portals, or participate in social support networks [69,70]. In addition, among older adults, those with higher cognitive skills are more likely to seek health information on the web than those with lower cognitive skills [68]. Other studies indicate an increase in internet use in older age groups; however, age was still not considered a predictor of HISB [70]. Although older adults may be heavy users of health services owing to increasing age-related and comorbid illnesses, they tend to be the lowest category of internet and other web-based health service users [71,72]. While these findings are specific to adults in Georgia, they are similar to other findings across the United States and the globe.

Social media allows users to quickly create and share content and participate in broad information sharing and consumption; different theoretical models propose that individuals are looking for action-oriented information, assessment of risk perception and responses, and more broadly, general information gathering [73]. In our study, the use of social media was not predictive of HL outcomes such as HSI, being able to critically appraise health information, FHI, or UHI. In addition, social media was the least used source of health information for individuals with both high and low HL in Georgia. Thus, although social media is a widely used platform for information dissemination, we found that it is not a significant source of health information nor does it appear to be related to HL outcomes. Some studies on information seeking of COVID-19 information, indicate that social media exposure may result in a significant overload of information that could lead to information anxiety and avoidance, thus having a negative impact on both HISB and HL outcomes [74]. Although this study was conducted during the pandemic, it did not focus on COVID-19 health information; rather, questions were asked about general HISBs.

Using cluster analysis, we were able to ascertain a high HISB and a low HISB cluster (39% and 61% of the sample, respectively). The high HISB cluster used all 5 sources of health information significantly more than the low HISB cluster in all HISB categories, and social media was used the least by both clusters. The high and low clusters were not differentiated by sex, educational level, county, or age. Interestingly, both clusters used social media the least as a health information source. Wang et al [75] posit that although social media networks are widely used and may facilitate HISB, they are also the perfect environment for spreading rumors and accurate information, and it is difficult for social media users to ascertain between the two. The lack of control over who can post information on the web has placed additional difficulty on discerning accurate scientific data from misinformation [20]. Social media content also changes quickly; users' cognitive limits may be maximized, which can lead to information overload, vulnerability,

uncertainty, and self-isolation [76]. Therefore, individuals may avoid social media channels when searching for health information.

The high HISB cluster exhibited higher HL across all 4 scales (HSI, CA of health information, FHI, and UHI). This is consistent with prior studies that indicate that having higher HL may influence a preference for information seeking over and above demographic variables [77,78]. This may not be causal; that is, individuals who seek health information may improve their HL owing to motivation for or better access to information in the same way that individuals with higher HL are more confident in seeking health information [77-79]. Studies indicate that patients who have higher HL may also have better patient engagement, have high self-advocacy, participate more in shared decision-making, and have better health outcomes [80-82]. Those patients with low HL may not have the capacity to seek health information from multiple sources because of their lower socioeconomic status, language barriers, and educational differences. Often, as compared with those with higher HL, those with lower HL are more likely to rely on health care providers' recommendations for their clinical course of action without seeking further information, signifying overlap between HISB and HL in socio-cognitive predictors such as perceived self-efficacy to obtain health information. In these cases, strategies should be implemented to increase patients' motivation to be informed on how to access, understand, and use health information from other sources. Health education practices targeting these populations may facilitate a greater understanding of clinical information and lead to healthier clinical outcomes [77,83,84].

Limitations

While this study sample mirrored the demographics of the state, we were only able to reach individuals who have computer access. Thus, we have reported findings only for individuals who have digital access and at least a minimum of digital literacy skills. As the recruitment was performed using web-based channels, sampling bias is a potential limitation of this study, as those who had difficulties in using these channels could be excluded from recruitment. Another limitation is that we were only able to survey participants in 1 southern US state. We stratified the sample to match the statewide demographic characteristics of geography and race but learned after data collection that sex and age distributions are largely skewed. Future studies should construct more complex stratification to account for this skewness in the data. Although we believe the findings are generalizable among Georgia residents, they may not be generalizable across other states.

Conclusions

Age and sex were significantly associated with HISBs. As older adults are more likely to use health services, they may benefit from having web-based resources to update them on their health status in real time and to provide accessible social support networks. Thus, there is a need to improve HISB skills of and interventions for older adults. Higher levels of HL are associated with greater HISB. Those with lower levels of HL may benefit from targeted strategies to improve their understanding of health information and how to access, understand, and use it, as greater understanding of health information is associated with healthier clinical outcomes. Further studies are needed, specifically those focused on HL, urbanicity, and access to health information.

Conflicts of Interest

None declared.

References

1. Lambert SD, Loiselle CG. Health information seeking behavior. Qual Health Res 2007 Oct;17(8):1006-1019. [doi: 10.1177/1049732307305199] [Medline: 17928475]
2. Faiman B, Tariman JD. Shared decision making: improving patient outcomes by understanding the benefits of and barriers to effective communication. Clin J Oncol Nurs 2019 Oct 01;23(5):540-542. [doi: 10.1188/19.CJON.540-542] [Medline: 31538972]
3. Peimani M, Nasli-Esfahani E, Sadeghi R. Patients' perceptions of patient-provider communication and diabetes care: a systematic review of quantitative and qualitative studies. Chronic Illn 2020 Mar;16(1):3-22. [doi: 10.1177/1742395318782378] [Medline: 29895167]
4. Zimmerman MS, Shaw Jr G. Health information seeking behaviour: a concept analysis. Health Info Libr J 2020 Sep;37(3):173-191. [doi: 10.1111/hir.12287] [Medline: 32052549]
5. Mirzaei A, Aslani P, Luca EJ, Schneider CR. Predictors of health information-seeking behavior: systematic literature review and network analysis. J Med Internet Res 2021 Jul 02;23(7):e21680 [FREE Full text] [doi: 10.2196/21680] [Medline: 33979776]
6. Yang Q, Chen Y, Wendorf Muhamad J. Social support, trust in health information, and health information-seeking behaviors (HISBs): a study using the 2012 Annenberg National Health Communication Survey (ANHCS). Health Commun 2017 Sep;32(9):1142-1150. [doi: 10.1080/10410236.2016.1214220] [Medline: 27589249]
7. Mendes Á, Abreu L, Vilar-Correia MR, Borlido-Santos J. "That should be left to doctors, that's what they are there for!"-exploring the reflexivity and trust of young adults when seeking health information. Health Commun 2017 Sep;32(9):1076-1081. [doi: 10.1080/10410236.2016.1199081] [Medline: 27485762]

8. Baker R, Mainous 3rd AG, Gray DP, Love MM. Exploration of the relationship between continuity, trust in regular doctors and patient satisfaction with consultations with family doctors. Scand J Prim Health Care 2003 Mar;21(1):27-32. [doi: 10.1080/0283430310000528] [Medline: 12718457]
9. Bundorf MK, Wagner TH, Singer SJ, Baker LC. Who searches the Internet for health information? Health Serv Res 2006 Jun;41(3 Pt 1):819-836 [FREE Full text] [doi: 10.1111/j.1475-6773.2006.00510.x] [Medline: 16704514]
10. Basch CH, MacLean SA, Romero RA, Ethan D. Health information seeking behavior among college students. J Community Health 2018 Dec;43(6):1094-1099. [doi: 10.1007/s10900-018-0526-9] [Medline: 29779074]
11. Zapata AM, Beaudreau SA, O'Hara R, Bereknyei Merrell S, Bruce J, Garrison-Diehn C, et al. Information-seeking about anxiety and perceptions about technology to teach coping skills in older veterans. Clin Gerontol 2018;41(4):346-356. [doi: 10.1080/07317115.2017.1359716] [Medline: 28967837]
12. Curtis BL, Ashford RD, Magnuson KI, Ryan-Pettes SR. Comparison of smartphone ownership, social media use, and willingness to use digital interventions between generation Z and millennials in the treatment of substance use: cross-sectional questionnaire study. J Med Internet Res 2019 Apr 17;21(4):e13050 [FREE Full text] [doi: 10.2196/13050] [Medline: 30994464]
13. Barker V. A generational comparison of social networking site use: the influence of age and social identity. Int J Aging Hum Dev 2012;74(2):163-187. [doi: 10.2190/AG.74.2.d] [Medline: 22808625]
14. Chaudhuri S, Le T, White C, Thompson H, Demiris G. Examining health information-seeking behaviors of older adults. Comput Inform Nurs 2013 Nov;31(11):547-553 [FREE Full text] [doi: 10.1097/01.NCN.0000432131.92020.42] [Medline: 23974574]
15. Feinberg I, Frijters J, Johnson-Lawrence V, Greenberg D, Nightingale E, Moodie C. Examining associations between health information seeking behavior and adult education status in the U.S.: an analysis of the 2012 PIAAC data. PLoS One 2016 Feb 16;11(2):e0148751 [FREE Full text] [doi: 10.1371/journal.pone.0148751] [Medline: 26882339]
16. Shah S, Diwan S, Kohan L, Rosenblum D, Gharibo C, Soin A, et al. The technological impact of COVID-19 on the future of education and health care delivery. Pain Physician 2020 Aug;23(4S):S367-S380 [FREE Full text] [Medline: 32942794]
17. Ramsey I, Corsini N, Peters MD, Eckert M. A rapid review of consumer health information needs and preferences. Patient Educ Couns 2017 Sep;100(9):1634-1642. [doi: 10.1016/j.pec.2017.04.005] [Medline: 28442155]
18. Alberti TL, Crawford SL. Health information-seeking behaviors and adherence to urgent care discharge instructions. J Am Assoc Nurse Pract 2020 Jun;32(6):438-446. [doi: 10.1097/JXX.0000000000000262] [Medline: 31651586]
19. Kortum P, Edwards C, Richards-Kortum R. The impact of inaccurate Internet health information in a secondary school learning environment. J Med Internet Res 2008 Jun 30;10(2):e17 [FREE Full text] [doi: 10.2196/jmir.986] [Medline: 18653441]
20. Rahnavardi M, Arabi MS, Ardalan G, Zamani N, Jahanbin M, Sohani F, et al. Accuracy and coverage of reproductive health information on the Internet accessed in English and Persian from Iran. J Fam Plann Reprod Health Care 2008 Jul;34(3):153-157 [FREE Full text] [doi: 10.1783/147118908784734882] [Medline: 18577313]
21. Khaleel I, Wimmer BC, Peterson GM, Zaidi ST, Roehrer E, Cummings E, et al. Health information overload among health consumers: a scoping review. Patient Educ Couns 2020 Jan;103(1):15-32. [doi: 10.1016/j.pec.2019.08.008] [Medline: 31451363]
22. Calleja N, AbdAllah A, Abad N, Ahmed N, Albarracin D, Altieri E, et al. A public health research agenda for managing infodemics: methods and results of the first WHO infodemiology conference. JMIR Infodemiology 2021 Sep 15;1(1):e30979 [FREE Full text] [doi: 10.2196/30979] [Medline: 34604708]
23. Vosoughi S, Roy D, Aral S. The spread of true and false news online. Science 2018 Mar 09;359(6380):1146-1151. [doi: 10.1126/science.aap9559] [Medline: 29590045]
24. Tasnim S, Hossain MM, Mazumder H. Impact of rumors and misinformation on COVID-19 in social media. J Prev Med Public Health 2020 May;53(3):171-174 [FREE Full text] [doi: 10.3961/jpmph.20.094] [Medline: 32498140]
25. Xie B, He D, Mercer T, Wang Y, Wu D, Fleischmann KR, et al. Global health crises are also information crises: a call to action. J Assoc Inf Sci Technol (forthcoming) 2020 Mar 13 [FREE Full text] [doi: 10.1002/asi.24357] [Medline: 32427189]
26. Crawford A, Serhal E. Digital health equity and COVID-19: the innovation curve cannot reinforce the social gradient of health. J Med Internet Res 2020 Jun 02;22(6):e19361 [FREE Full text] [doi: 10.2196/19361] [Medline: 32452816]
27. Geller AI, Budnitz DS, Dubendris H, Gharpure R, Soe M, Wu H, et al. Surveillance of COVID-19 vaccination in nursing homes, United States, December 2020-July 2021. Public Health Rep 2022;137(2):239-243. [doi: 10.1177/00333549211066168] [Medline: 35125027]
28. Nutbeam D. The evolving concept of health literacy. Soc Sci Med 2008 Dec;67(12):2072-2078. [doi: 10.1016/j.socscimed.2008.09.050] [Medline: 18952344]
29. Sørensen K, Van den Broucke S, Fullam J, Doyle G, Pelikan J, Slonska Z, (HLS-EU) Consortium Health Literacy Project European. Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health 2012 Jan 25;12:80 [FREE Full text] [doi: 10.1186/1471-2458-12-80] [Medline: 22276600]
30. Helitzer D, Hollis C, Sanders M, Roybal S. Addressing the "other" health literacy competencies--knowledge, dispositions, and oral/aural communication: development of TALKDOC, an intervention assessment tool. J Health Commun 2012;17 Suppl 3:160-175 [FREE Full text] [doi: 10.1080/10810730.2012.712613] [Medline: 23030568]

31. Sentell T, Vamos S, Okan O. Interdisciplinary perspectives on health literacy research around the world: more important than ever in a time of COVID-19. Int J Environ Res Public Health 2020 Apr 26;17(9):3010 [FREE Full text] [doi: 10.3390/ijerph17093010] [Medline: 32357457]
32. Velardo S. The nuances of health literacy, nutrition literacy, and food literacy. J Nutr Educ Behav 2015;47(4):385-9.e1. [doi: 10.1016/j.jneb.2015.04.328] [Medline: 26026651]
33. Fairbrother H, Curtis P, Goyder E. Making health information meaningful: children's health literacy practices. SSM Popul Health 2016 Jul 16;2:476-484 [FREE Full text] [doi: 10.1016/j.ssmph.2016.06.005] [Medline: 29349163]
34. Smith SK, Dixon A, Trevena L, Nutbeam D, McCaffery KJ. Exploring patient involvement in healthcare decision making across different education and functional health literacy groups. Soc Sci Med 2009 Dec;69(12):1805-1812. [doi: 10.1016/j.socscimed.2009.09.056] [Medline: 19846245]
35. Chin J, Morrow DG, Stine-Morrow EA, Conner-Garcia T, Graumlich JF, Murray MD. The process-knowledge model of health literacy: evidence from a componential analysis of two commonly used measures. J Health Commun 2011;16 Suppl 3:222-241 [FREE Full text] [doi: 10.1080/10810730.2011.604702] [Medline: 21951254]
36. Manganello JA. Health literacy and adolescents: a framework and agenda for future research. Health Educ Res 2008 Oct;23(5):840-847. [doi: 10.1093/her/cym069] [Medline: 18024979]
37. Zarcadoolas C, Pleasant A, Greer DS. Understanding health literacy: an expanded model. Health Promot Int 2005 Jun;20(2):195-203. [doi: 10.1093/heapro/dah609] [Medline: 15788526]
38. Norgaard O, Furstrand D, Klokker L, Kanoe A, Batterham R, Kayser L, et al. The e-health literacy framework: a conceptual framework for characterizing e-health users and their interaction with e-health systems. Knowl Manag E-Learn 2015;7(4):522-540. [doi: 10.34105/j.kmel.2015.07.035]
39. Chesser A, Burke A, Reyes J, Rohrberg T. Navigating the digital divide: a systematic review of eHealth literacy in underserved populations in the United States. Inform Health Soc Care 2016;41(1):1-19. [doi: 10.3109/17538157.2014.948171] [Medline: 25710808]
40. Levin-Zamir D, Bertschi I. Media health literacy, eHealth literacy, and the role of the social environment in context. Int J Environ Res Public Health 2018 Aug 03;15(8):1643 [FREE Full text] [doi: 10.3390/ijerph15081643] [Medline: 30081465]
41. Wang C, Wu X, Qi H. A comprehensive analysis of e-health literacy research focuses and trends. Healthcare (Basel) 2021 Dec 30;10(1):66 [FREE Full text] [doi: 10.3390/healthcare10010066] [Medline: 35052230]
42. Adams A, Adams R, Thorogood M, Buckingham C. Barriers to the use of e-health technology in nurse practitioner-patient consultations. Inform Prim Care 2007;15(2):103-109 [FREE Full text] [doi: 10.14236/jhi.v15i2.648] [Medline: 17877872]
43. Osborne RH, Batterham RW, Elsworth GR, Hawkins M, Buchbinder R. The grounded psychometric development and initial validation of the Health Literacy Questionnaire (HLQ). BMC Public Health 2013 Jul 16;13:658 [FREE Full text] [doi: 10.1186/1471-2458-13-658] [Medline: 23855504]
44. Hawkins M, Gill SD, Batterham R, Elsworth GR, Osborne RH. The Health Literacy Questionnaire (HLQ) at the patient-clinician interface: a qualitative study of what patients and clinicians mean by their HLQ scores. BMC Health Serv Res 2017 Apr 27;17(1):309 [FREE Full text] [doi: 10.1186/s12913-017-2254-8] [Medline: 28449680]
45. Morris RL, Soh SE, Hill KD, Buchbinder R, Lowthian JA, Redfern J, et al. Measurement properties of the Health Literacy Questionnaire (HLQ) among older adults who present to the emergency department after a fall: a Rasch analysis. BMC Health Serv Res 2017 Aug 29;17(1):605 [FREE Full text] [doi: 10.1186/s12913-017-2520-9] [Medline: 28851344]
46. Nolte S, Osborne RH, Dwinger S, Elsworth GR, Conrad ML, Rose M, et al. German translation, cultural adaptation, and validation of the Health Literacy Questionnaire (HLQ). PLoS One 2017 Feb 24;12(2):e0172340 [FREE Full text] [doi: 10.1371/journal.pone.0172340] [Medline: 28234987]
47. Farmanova E, Bonneville L, Bouchard L. Organizational health literacy: review of theories, frameworks, guides, and implementation issues. Inquiry 2018;55:46958018757848 [FREE Full text] [doi: 10.1177/0046958018757848] [Medline: 29569968]
48. Weaver NL, Wray RJ, Zellin S, Gautam K, Jupka K. Advancing organizational health literacy in health care organizations serving high-needs populations: a case study. J Health Commun 2012;17 Suppl 3:55-66. [doi: 10.1080/10810730.2012.714442] [Medline: 23030561]
49. Palumbo R. Leveraging organizational health literacy to enhance health promotion and risk prevention: a narrative and interpretive literature review. Yale J Biol Med 2021 Mar;94(1):115-128 [FREE Full text] [Medline: 33795988]
50. Nijman J, Hendriks M, Brabers A, de Jong J, Rademakers J. Patient activation and health literacy as predictors of health information use in a general sample of Dutch health care consumers. J Health Commun 2014;19(8):955-969. [doi: 10.1080/10810730.2013.837561] [Medline: 24397280]
51. Kim YC, Lim JY, Park K. Effects of health literacy and social capital on health information behavior. J Health Commun 2015;20(9):1084-1094. [doi: 10.1080/10810730.2015.1018636] [Medline: 26166008]
52. Rouquette A, Nadot T, Labitrie P, Van den Broucke S, Mancini J, Rigal L, et al. Validity and measurement invariance across sex, age, and education level of the French short versions of the European Health Literacy Survey Questionnaire. PLoS One 2018 Dec 6;13(12):e0208091 [FREE Full text] [doi: 10.1371/journal.pone.0208091] [Medline: 30521552]
53. Baker DW, Wolf MS, Feinglass J, Thompson JA, Gazmararian JA, Huang J. Health literacy and mortality among elderly persons. Arch Intern Med 2007 Jul 23;167(14):1503-1509. [doi: 10.1001/archinte.167.14.1503] [Medline: 17646604]

54. Matsuyama RK, Wilson-Genderson M, Kuhn L, Moghanaki D, Vachhani H, Paasche-Orlow M. Education level, not health literacy, associated with information needs for patients with cancer. Patient Educ Couns 2011 Dec;85(3):e229-e236. [doi: 10.1016/j.pec.2011.03.022] [Medline: 21543182]
55. Halverson J, Martinez-Donate A, Trentham-Dietz A, Walsh MC, Strickland JS, Palta M, et al. Health literacy and urbanicity among cancer patients. J Rural Health 2013;29(4):392-402 [FREE Full text] [doi: 10.1111/jrh.12018] [Medline: 24088213]
56. Waverijn G, Heijmans M, Spreeuwenberg P, Groenewegen PP. Associations between neighborhood social capital, health literacy, and self-rated health among people with chronic illness. J Health Commun 2016;21(sup2):36-44. [doi: 10.1080/10810730.2016.1179369] [Medline: 27548376]
57. Wong PK, Christie L, Johnston J, Bowling A, Freeman D, Joshua F, et al. How well do patients understand written instructions?: health literacy assessment in rural and urban rheumatology outpatients. Medicine (Baltimore) 2014 Nov;93(25):e129 [FREE Full text] [doi: 10.1097/MD.0000000000000129] [Medline: 25437024]
58. Wang W, Zhang Y, Lin B, Mei Y, Ping Z, Zhang Z. The urban-rural disparity in the status and risk factors of health literacy: a cross-sectional survey in central China. Int J Environ Res Public Health 2020 May 29;17(11):3848 [FREE Full text] [doi: 10.3390/ijerph17113848] [Medline: 32485790]
59. Chen X, Orom H, Hay JL, Waters EA, Schofield E, Li Y, et al. Differences in rural and urban health information access and use. J Rural Health 2019 Jun;35(3):405-417 [FREE Full text] [doi: 10.1111/jrh.12335] [Medline: 30444935]
60. Miller CE, Vasan RS. The southern rural health and mortality penalty: a review of regional health inequities in the United States. Soc Sci Med 2021 Jan;268:113443 [FREE Full text] [doi: 10.1016/j.socscimed.2020.113443] [Medline: 33137680]
61. Holahan J, Buettgens M, Dorn S. The Cost of Not Expanding Medicaid. Kaiser Family Foundation. 2013 Jul 17. URL: https://www.kff.org/medicaid/report/the-cost-of-not-expanding-medicaid/ [accessed 2022-03-21]
62. America's Health Rankings 2021 Annual Report. United Health Foundation. 2021. URL: https://www.americashealthrankings.org/learn/reports/2021-annual-report [accessed 2022-03-21]
63. Decker SL, Kenney GM, Long SK. Characteristics of uninsured low-income adults in states expanding vs not expanding Medicaid. JAMA Intern Med 2014 Jun;174(6):988-989 [FREE Full text] [doi: 10.1001/jamainternmed.2014.518] [Medline: 24710741]
64. United States Census Bureau. URL: https://data.census.gov/cedsci/map?q=B03002&g=0400000US13&tid=ACSDT5Y2020.B03002&cid=B03002_001E&vintage=2020&mode=thematic&loc=38.8800,-98.0000,z3.0000 [accessed 2022-06-09]
65. Yoon H, Jang Y, Kim S, Speasmaker A, Nam I. Trends in internet use among older adults in the United States, 2011-2016. J Appl Gerontol 2021 May;40(5):466-470. [doi: 10.1177/0733464820908427] [Medline: 32131670]
66. Rains SA. Perceptions of traditional information sources and use of the world wide web to seek health information: findings from the health information national trends survey. J Health Commun 2007;12(7):667-680. [doi: 10.1080/10810730701619992] [Medline: 17934943]
67. Rice RE. Influences, usage, and outcomes of Internet health information searching: multivariate results from the Pew surveys. Int J Med Inform 2006 Jan;75(1):8-28. [doi: 10.1016/j.ijmedinf.2005.07.032] [Medline: 16125453]
68. Percheski C, Hargittai E. Health information-seeking in the digital age. J Am Coll Health 2011;59(5):379-386. [doi: 10.1080/07448481.2010.513406] [Medline: 21500056]
69. Cosco TD, Firth J, Vahia I, Sixsmith A, Torous J. Mobilizing mHealth data collection in older adults: challenges and opportunities. JMIR Aging 2019 Mar 19;2(1):e10019 [FREE Full text] [doi: 10.2196/10019] [Medline: 31518253]
70. Huerta TR, Walker DM, Johnson T, Ford EW. A time series analysis of cancer-related information seeking: hints from the Health Information National Trends Survey (HINTS) 2003-2014. J Health Commun 2016 Sep;21(9):1031-1038. [doi: 10.1080/10810730.2016.1204381] [Medline: 27565190]
71. Theis S, Schäfer K, Schäfer D, Jochems N, Nitsch V, Mertens A. The relationship between individual coping and the need to have and seek health information among older adults: exploratory mixed methods study. JMIR Hum Factors 2021 Feb 01;8(1):e15858 [FREE Full text] [doi: 10.2196/15858] [Medline: 33522973]
72. Mizrachi Y, Shahrabani S, Nachmani M, Hornik A. Obstacles to using online health services among adults age 50 and up and the role of family support in overcoming them. Isr J Health Policy Res 2020 Aug 21;9(1):42 [FREE Full text] [doi: 10.1186/s13584-020-00398-x] [Medline: 32825840]
73. Wang X, Shi J, Kong H. Online health information seeking: a review and meta-analysis. Health Commun 2021 Sep;36(10):1163-1175. [doi: 10.1080/10410236.2020.1748829] [Medline: 32290679]
74. Soroya SH, Farooq A, Mahmood K, Isoaho J, Zara SE. From information seeking to information avoidance: understanding the health information behavior during a global health crisis. Inf Process Manag 2021 Mar;58(2):102440 [FREE Full text] [doi: 10.1016/j.ipm.2020.102440] [Medline: 33281273]
75. Wang Y, McKee M, Torbica A, Stuckler D. Systematic literature review on the spread of health-related misinformation on social media. Soc Sci Med 2019 Nov;240:112552 [FREE Full text] [doi: 10.1016/j.socscimed.2019.112552] [Medline: 31561111]
76. Farooq A, Laato S, Islam AK. Impact of online information on self-isolation intention during the COVID-19 pandemic: cross-sectional study. J Med Internet Res 2020 May 06;22(5):e19128 [FREE Full text] [doi: 10.2196/19128] [Medline: 32330115]

77. Kim SH, Utz S. Association of health literacy with health information-seeking preference in older people: a correlational, descriptive study. Nurs Health Sci 2018 Sep;20(3):355-360. [doi: 10.1111/nhs.12413] [Medline: 29488295]
78. Cashin A, Heartfield M, Cox D, Dunn S, Stasa H. Knowledge and motivation: two elements of health literacy that remain low with regard to nurse practitioners in Australia. Aust Health Rev 2015 Sep;39(4):470-475. [doi: 10.1071/AH14126] [Medline: 25751752]
79. Kim SH, Lee A. Health-literacy-sensitive diabetes self-management interventions: a systematic review and meta-analysis. Worldviews Evid Based Nurs 2016 Aug;13(4):324-333. [doi: 10.1111/wvn.12157] [Medline: 27104337]
80. Smith SG, Curtis LM, Wardle J, von Wagner C, Wolf MS. Skill set or mind set? Associations between health literacy, patient activation and health. PLoS One 2013 Sep 4;8(9):e74373 [FREE Full text] [doi: 10.1371/journal.pone.0074373] [Medline: 24023942]
81. Martin LT, Schonlau M, Haas A, Derose KP, Rosenfeld L, Buka SL, et al. Patient activation and advocacy: which literacy skills matter most? J Health Commun 2011;16 Suppl 3:177-190 [FREE Full text] [doi: 10.1080/10810730.2011.604705] [Medline: 21951251]
82. Hibbard JH, Greene J. What the evidence shows about patient activation: better health outcomes and care experiences; fewer data on costs. Health Aff (Millwood) 2013 Feb;32(2):207-214. [doi: 10.1377/hlthaff.2012.1061] [Medline: 23381511]
83. Matterne U, Egger N, Tempes J, Tischer C, Lander J, Dierks M, et al. Health literacy in the general population in the context of epidemic or pandemic coronavirus outbreak situations: rapid scoping review. Patient Educ Couns 2021 Feb;104(2):223-234 [FREE Full text] [doi: 10.1016/j.pec.2020.10.012] [Medline: 33109429]
84. Ellis J, Mullan J, Worsley A, Pai N. The role of health literacy and social networks in arthritis patients' health information-seeking behavior: a qualitative study. Int J Family Med 2012;2012:397039 [FREE Full text] [doi: 10.1155/2012/397039] [Medline: 22997575]

Abbreviations

CA: critical appraisal
FHI: finding health information
HISB: health information–seeking behavior
HL: health literacy
HLQ: Health Literacy Questionnaire
HSI: have sufficient information
OHL: organizational health literacy
UHI: understanding health information

Edited by A Mavragani; submitted 04.11.21; peer-reviewed by D He, S Song; comments to author 01.03.22; revised version received 28.03.22; accepted 13.04.22; published 15.06.22.

Please cite as:
Rao N, Tighe EL, Feinberg I
The Dispersion of Health Information–Seeking Behavior and Health Literacy in a State in the Southern United States: Cross-sectional Study
JMIR Form Res 2022;6(6):e34708
URL: https://formative.jmir.org/2022/6/e34708
doi:10.2196/34708
PMID:35704357

Original Paper

Assessing the Initial Validity of the PortionSize App to Estimate Dietary Intake Among Adults: Pilot and Feasibility App Validation Study

Sanjoy Saha[1], PhD; Chloe Panizza Lozano[1], PhD; Stephanie Broyles[1], PhD; Corby K Martin[1], PhD; John W Apolzan[1], PhD

Pennington Biomedical Research Center, Louisiana State University System, Baton Rouge, LA, United States

Corresponding Author:
John W Apolzan, PhD
Pennington Biomedical Research Center
Louisiana State University System
6400 Perkins Road
Baton Rouge, LA, 70808
United States
Phone: 1 2257632827
Email: john.apolzan@pbrc.edu

Abstract

Background: Accurately assessing dietary intake can promote improved nutrition. The PortionSize app (Pennington Biomedical Research Center) was designed to quantify and provide real-time feedback on the intake of energy, food groups, saturated fat, and added sugar.

Objective: This study aimed to assess the preliminary feasibility and validity of estimating food intake via the PortionSize app among adults.

Methods: A total of 15 adults (aged 18-65 years) were recruited and trained to quantify the food intake from a simulated meal by using PortionSize. Trained personnel prepared 15 simulated meals and covertly weighed (weigh back) the amount of food provided to participants as well as food waste. Equivalence tests (±25% bounds) were performed to compare PortionSize to the weigh back method.

Results: Participants were aged a mean of 28 (SD 12) years, and 11 were female. The mean energy intake estimated with PortionSize was 742.9 (SD 328.2) kcal, and that estimated via weigh back was 659.3 (SD 190.7) kcal (energy intake difference: mean 83.5, SD 287.5 kcal). The methods were not equivalent in estimating energy intake (P=.18), and PortionSize overestimated energy intake by 83.5 kcal (12.7%) at the meal level. Estimates of portion sizes (gram weight; P=.01), total sugar (P=.049), fruit servings (P=.01), and dairy servings (P=.047) from PortionSize were equivalent to those estimated via weigh back. PortionSize was not equivalent to weigh back with regard to estimates for carbohydrate (P=.10), fat (P=.32), vegetable (P=.37), grain (P=.31), and protein servings (P=.87).

Conclusions: Due to power limitations, the equivalence tests had large equivalence bounds. Though preliminary, the results of this small pilot study warrant the further adaptation, development, and validation of PortionSize as a means to estimate energy intake and provide users with real-time and actionable dietary feedback.

(JMIR Form Res 2022;6(6):e38283) doi:10.2196/38283

KEYWORDS

dietary assessment; eating; food intake; energy intake; portion size; mHealth; digital health; eHealth; nutrition; food groups

Introduction

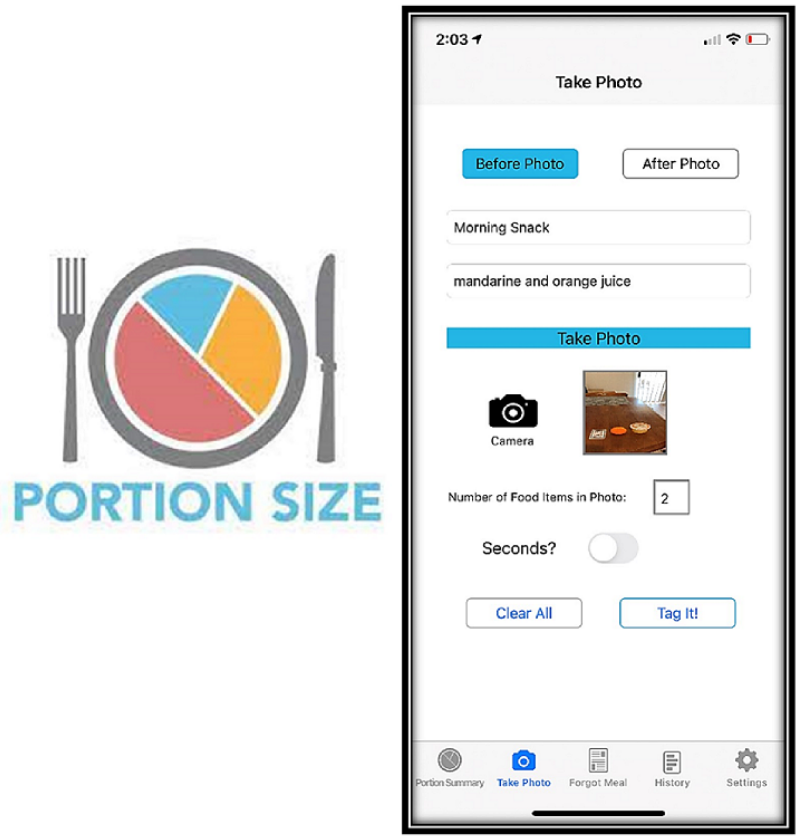
Accurately quantifying food intake is important to managing body weight; improving health and nutrition; and reducing the risk of chronic diseases, such as malnutrition, diabetes, and obesity [1-4]. Valid food intake assessment methods are needed to determine whether individual dietary patterns and nutrient intake meet the recommended levels [1-6]. Moreover, sufficiently accurate methods are needed that can provide people with information in real time about what foods they select and eat to facilitate the modification of dietary behaviors when they occur [1,7-10].

Traditional dietary intake assessment methods, including 24-hour dietary recall, food frequency questionnaires, and food records, have been widely used in nutritional research. These conventional methods have some advantages, as well as limitations. Recall-based methods rely on participants' memory to recall what foods were consumed, how they were prepared, and how much of each food they consumed (ie, participants must estimate portion size). These traditional methods of dietary assessment are also time consuming [1,10,11]. Advancements in technology allow for unique perspectives when estimating energy and nutrient intake [10]. Web and food photography or food image–based methods may reduce user burden and provide more accurate estimates of food intake [10,12]. Although some of these methods may be accurate [13-15], they also have limitations. For example, food photography–based methods, such as the remote food photography method (RFPM), require trained human raters to analyze food images and quantify food intake. Consequently, such methods do not provide immediate feedback about food intake to the users, are not scalable, and have little to no cost advantage over more traditional methods [3,13,16].

PortionSize (Pennington Biomedical Research Center [PBRC]) is a newly developed smartphone app and method for estimating food intake that also relies on images of foods. Rather than human raters estimating portion size based on food images, the app integrates templates and other techniques that allow users to estimate portion sizes in real time [7]. Consequently, the users receive immediate feedback about their food selections prior to eating, which theoretically allows users to modify their food selections to better adhere to certain energy intake levels, food group recommendations, or macronutrient levels. The users also receive information about their food intake after they eat. The information provided to the users includes energy intake; fruit, vegetable, grain, protein, and dairy servings; and amounts of saturated fat and added sugar. The users also receive feedback about the extent to which their intake throughout the day is meeting specific energy intake and food group goals (eg, United States Department of Agriculture [USDA] MyPlate–recommended food groups: fruits, vegetables, grains, protein, and dairy). The results of the users estimating portion size and the ability of the users to receive food intake information immediately are expected reductions in validity and accuracy, particularly when compared to those of the RFPM [3,7,13]. To our knowledge, PortionSize is one of the first apps to provide users with food intake adherence data about their food selections before meals are consumed, after meals, and cumulatively throughout the day. We expect that the PortionSize app will help users overcome the limitations inherent with the RFPM.

The portion sizes of food have increased in the United States, and at fast-food chains, portions have increased by 2 to 5 times the original serving size [17]. Without visual aids however, it is very difficult for people [18], including trained registered dietitians [19], to accurately estimate portion sizes. Nevertheless, most dietary intake assessment methods focus on energy and nutrient intake and are not able to capture information on whether consumers are meeting the USDA MyPlate–recommended daily servings [20]. Therefore, there remains a significant need for methods that are sufficiently accurate to provide researchers with good outcome data and guide health promotion efforts while remaining scalable and affordable. PortionSize relies on emerging technology (eg, augmented reality) to improve accuracy and minimize the amount of missing data. PortionSize integrates ecological momentary assessment methods [21] to drive data completeness and quality (Figure 1). Validation studies of food intake assessment methods play an important role in identifying key areas to improve the accuracy of food intake estimation [18]. To collect preliminary data, assess initial validity, and identify areas of improvement for the PortionSize app, we conducted this pilot study. The aim of this pilot and feasibility study was to collect preliminary validity data on food intake that are estimated with the PortionSize app and compare them with data on weighed food. The secondary aim was to explore participants' perceived satisfaction with the PortionSize app.

Figure 1. The PortionSize app allows users to take before-meal photos and after-meal photos.



Methods

Ethics Approval

This study was conducted in accordance with the Declaration of Helsinki [22], and all procedures involving human subjects were approved by the institutional review board (IRB) at the PBRC (IRB Federal Wide Assurance number: 00006218). The trial was registered at ClinicalTrails.gov (trial number: NCT04494971) prior to recruiting participants in this study. Written informed consent was obtained from all subjects.

Recruitment and Participants

In this pilot study, 15 adult participants were enrolled, following the recommendation made by Hertzog [23] that 10 to 15 participants are sufficient for a pilot study. Advertisements on the PBRC Current Research Trials web page and the PBRC Facebook pages were used to recruit participants. We also distributed flyers at Louisiana State University. Participants who were interested in this study emailed the study team and then completed a phone screen. Preliminary eligible participants were scheduled for an in-person screening visit. Those who qualified and remained interested enrolled in the study. The eligibility criteria were adults aged 18 to 65 years and a BMI within the range of 18.5 to 45 kg/m^2. Participants who reported an eating disorder or serious mental illness, pregnant women, and breastfeeding mothers were excluded from this study. The aim of this pilot and feasibility study was to assess the PortionSize app's performance when participants ate typical meals; hence, these exclusion criteria eliminated participants whose eating patterns or meal sizes may have been atypical. Each participant was given a subject ID number to ensure confidentiality, and collected data were stored in the password-protected PBRC server. The app is Health Insurance Portability and Accountability Act compliant. Participants were compensated for their successful completion of the study.

Procedures

Participants completed a demographic questionnaire, and trained research staff conducted anthropometric measurements (height and weight) of the participants. Afterward, participants were trained to use the PortionSize app to measure food intake. During training, participants practiced assessing food intake by using food models. The entire session took about 1.5 hours.

Measures

Demographics and Anthropometrics

Participants' age, race, ethnicity, sex, marital status, education level, height, and weight were collected during their visit to the center. BMI was calculated from participants' objectively measured height and weight.

Directly Weighed Food Intake to Prepare Simulated Test Meals

Trained research staff prepared simulated test meals via direct observation. Such simulated meals served as the criterion measures of portion size, energy content, and food group

quantity for comparisons with the food intake assessments by participants using the PortionSize app. Participants' energy requirements were calculated by using sex-specific formulas [24]. Energy requirements were multiplied by 1.3, and the food selection for the simulated test meal included 30% of this value, which represents a typical lunch. The meals consisted of at least 3 food items and 1 calorie-containing beverage. Menus of meals were selected from a list of commonly consumed foods from a previous study [25]. Further, 3 participants were provided with the same simulated meal food menu (Table S1 in Multimedia Appendix 1); however, the meals differed in terms of portion size. Plate waste was determined at the individual food item level. This measure ranged from 0% and 100% and was right-skewed; as such, the mean plate waste was around 5% of the foods provided, which was similar to the actual plate waste from our free-living data (around 3%) [25]. Simulated food provision and plate waste were covertly weighed, and food intake was calculated by difference.

Food Intake Estimation Using the PortionSize App

Participants were instructed to use the PortionSize app to estimate food provision and waste. After the assessment, participants immediately obtained estimated feedback on their food intake, including energy intake, servings of different food groups (fruits, vegetables, grains, dairy, and protein), and amounts of selected nutrients (saturated fat and added sugar). The PortionSize app currently contains a database of around 1150 food items that are linked with the food codes in the Food and Nutrition Dataset for Dietary Studies (FNDDS) database [26]. Participants identified food items (from the served simulated meals) and associated food codes by selecting food items within the PortionSize app. A summary of details about the PortionSize app are included in the supplementary materials (Multimedia Appendix 1).

User Satisfaction Survey and the Computer System Usability Questionnaire

After completing the food intake assessments, participants completed 2 surveys. We adapted a 10-item user satisfaction survey that was administered in prior studies to quantify satisfaction, ease of use, and the adequacy of training for PortionSize [27,28]. The items were generated to obtain user satisfaction data and feedback about the app that could be used to identify areas where the app requires improvement. We did not rely on a formal framework when developing the survey. All items were rated on a scale ranging from 1 to 6, with 1 indicating "extremely dissatisfied," "very difficult," "not at all," or "not appropriate" and 6 indicating "extremely satisfied," "extremely easy," "very much," or "very appropriate."

Participants also completed the Computer System Usability Questionnaire (CSUQ; a 7-point rating scale)—a standardized, reliable, and valid questionnaire that was originally designed to evaluate computer programs [27,29]. It has been used to quantify the usability of mobile phone apps [30,31]. A rating of 7 represented "strongly disagree," and a rating of 1 indicated "strongly agree." The CSUQ provides an overall satisfaction score and scores for system usefulness, information quality, and interface quality [29].

Data Analysis

All statistical analyses were performed by using IBM SPSS software (version 28.0.1; IBM Corporation) and SAS/STAT software (version 9.4; SAS Institute Inc). The primary analysis was assessing the equivalence between the PortionSize app and the weigh back method by using equivalence tests, specifically the two one-sided *t* test method [32]. The primary outcome variable was the measured energy (kcal) calculated via the weigh back method at the meal level. The equivalence bounds were set at ±25%. These bounds are large, but they reflected the appropriate statistical power for a pilot study and were used in a similar pilot study [15]. A Bland-Altman analysis [33] was performed to test for differences in error variance over levels of the variable being measured (eg, food intake). We also calculated error from PortionSize in relation to the criterion measure (weigh back) by using 2-tailed dependent samples *t* tests to compare portion sizes and the intake of energy, food group servings (fruits, vegetables, grains, dairy, and protein), macronutrients (carbohydrates, fat, and protein), selected nutrients (saturated fat, cholesterol, dietary fiber, total sugar, and added sugar), and selected micronutrients (sodium, calcium, iron, potassium, and vitamin D). These results are presented in the supplementary materials (Multimedia Appendix 1). The inclusion of selected nutrients for analysis was determined based upon the nutrition facts panels. We estimated the mean percent difference (ie, [(PortionSize − weigh back)/weigh back] × 100) at the group level to avoid having a 0 value as a denominator for each nutrient. The significance level was set at .05. User satisfaction and CSUQ survey results were presented primarily as frequencies and percentages.

Results

Participants' Characteristics

A total of 21 participants completed the phone screening, and 15 participants were enrolled and completed the study. Of the 15 participants, 11 (73%) were female (Table 1). The mean age of the participants was 28 (SD 12) years, and the BMI (kg/m^2) range was 18.8 to 41.8 kg/m^2.

Table 1. Background characteristics of participants (N=15).

Variables	Value
Sex, n (%)	
Male	4 (27)
Female	11 (73)
Race and ethnicity, n (%)	
Black or African American	1 (7)
White	14 (93)
Education, n (%)	
High school diploma or General Educational Development	1 (7)
Some college	7 (47)
Bachelor's degree	5 (33)
Postgraduate degree	2 (13)
Employment, n (%)	
Unemployed	2 (13)
Full-time employment	4 (27)
Part-time employment	7 (47)
Retired	1 (7)
Other: student	1 (7)
Age (years), mean (SD; range)	28.0 (12.2; 20-57)
Height (cm), mean (SD; range)	168.1 (10.4; 147.3-182.9)
Weight (kg), mean (SD; range)	68.3 (19.8; 50.4-113.4)
BMI (kg/m^2), mean (SD; range)	24.1 (6.6; 18.8-41.8)

Estimation of Energy Intake

Table 2 indicates that the mean energy intake estimated with the PortionSize app (742.9, SD 328.2 kcal) was not equivalent (P=.18) to the mean estimated from the weighed meals (659.3, SD 190.7 kcal). The mean energy intake difference between the two methods was 83.5 (95% CI –480.0 to 647.0) kcal, and the mean percent error for the estimation of energy intake was 12.7% (Table 2 and Table S2 in Multimedia Appendix 1). PortionSize underestimated energy intake at lower levels of intake, but overestimation occurred and increased with higher levels of intake (Figure 2), as indicated by a significant regression equation (R^2=0.300; adjusted R^2=0.246; P=.03).

Table 2. Comparison of portion size, energy, and nutrient intake estimates between PortionSize and the weigh back method (meals: N=15).

	PortionSize app		Weigh back		Difference		Equivalence at ±25%, *P* value	Mean percent error[a]
	Mean	SD	Mean	SD	Mean	SD		
Energy (kcal)	742.9	328.2	659.3	190.7	83.5	287.5	.18	12.7
Portion size (g)	674.3	222.8	716.9	207.2	−42.7	303.9	.03[b]	−6
Total fruits (servings[c])	0.2	0.3	0.3	0.4	−0.1	0.4	.01[b]	−33.3
Total vegetables (servings[c])	0.6	0.3	0.6	0.4	0.0	0.2	.37	0
Total grains (servings[d])	1.7	1.7	1.2	1.1	0.5	0.8	.31	41.7
Total dairy (servings[c])	0.4	0.6	0.5	0.6	−0.1	0.7	.047[b]	−20
Total protein (servings[d])	3.1	3.6	2.8	2.9	0.3	2.0	.87	10.7
Saturated fat (g)	10.8	6.8	10.4	4.1	0.4	6.1	.11	3.8
Added sugar (teaspoons)	8.8	7.3	8.5	5.1	0.2	4.8	.14	2.4
Protein (g)	35.3	27.9	32.7	19.8	2.6	17.4	.25	8
Total fat (g)	32.5	22.4	27.1	10.8	5.4	18.7	.32	19.9
Carbohydrates (g)	78.9	49.6	72.0	28.4	6.9	33.8	.10	9.6
Dietary fiber (g)	4.6	3.2	4.1	1.3	0.5	3.6	.39	12.2
Total sugar (g)	45.6	27.6	46.5	15.7	−1.0	23.7	.049[b]	−2.2
Cholesterol (mg)	110.3	94.0	103.3	73.0	7.1	61.9	.52	6.9
Sodium (mg)	1200.8	562.0	940.5	376.7	260.3	449.5	.68	27.7
Calcium (mg)	214.3	185.4	254.7	176.5	−40.4	171.7	.09	−15.9
Iron (mg)	4.3	3.1	3.3	2.3	1.0	1.4	.72	30.3
Potassium (mg)	888.5	576.6	831.9	431.8	56.6	361.1	.049[b]	6.8
Vitamin D (µg)	0.8	1.9	1.5	2.0	−0.6	2.1	.62	−40

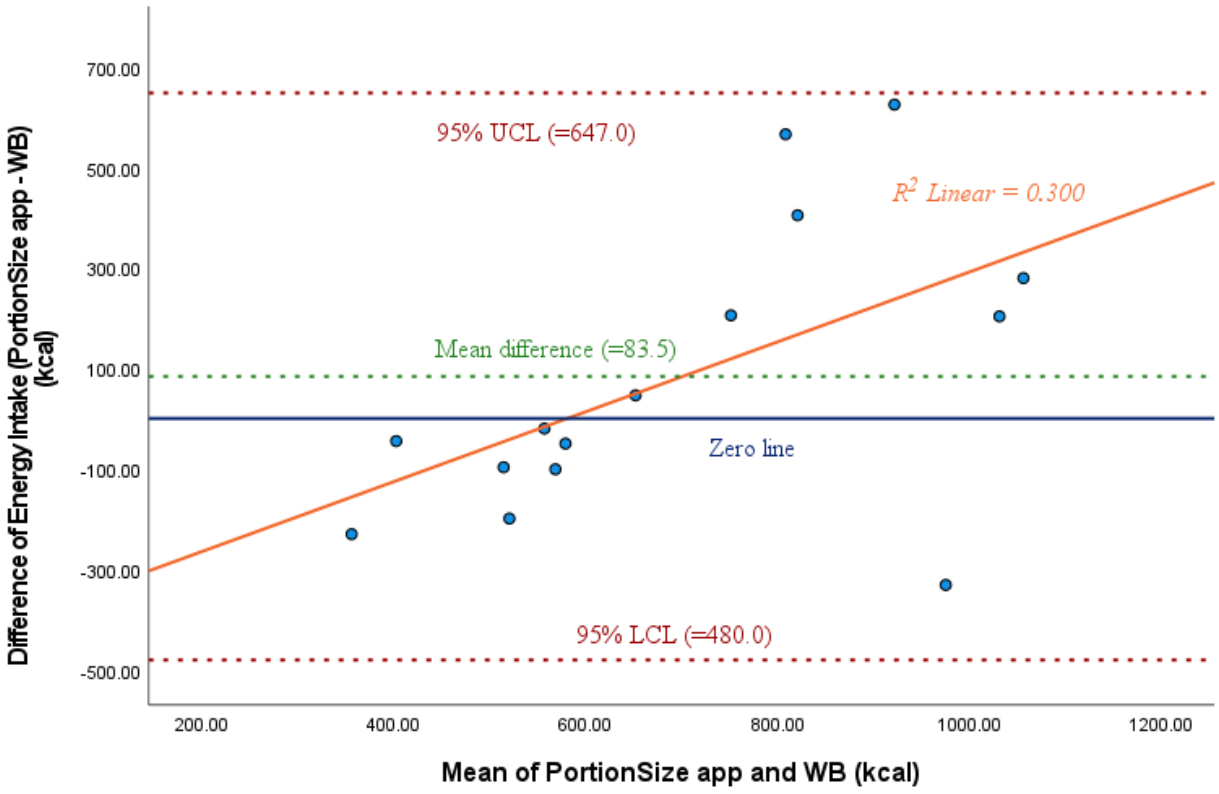
[a]Mean percent error = ([PortionSize – weigh back]/weigh back) × 100.

[b]Significant equivalence (level of significance at $P<.05$).

[c]Servings were cup equivalents.

[d]Servings were ounce equivalents.

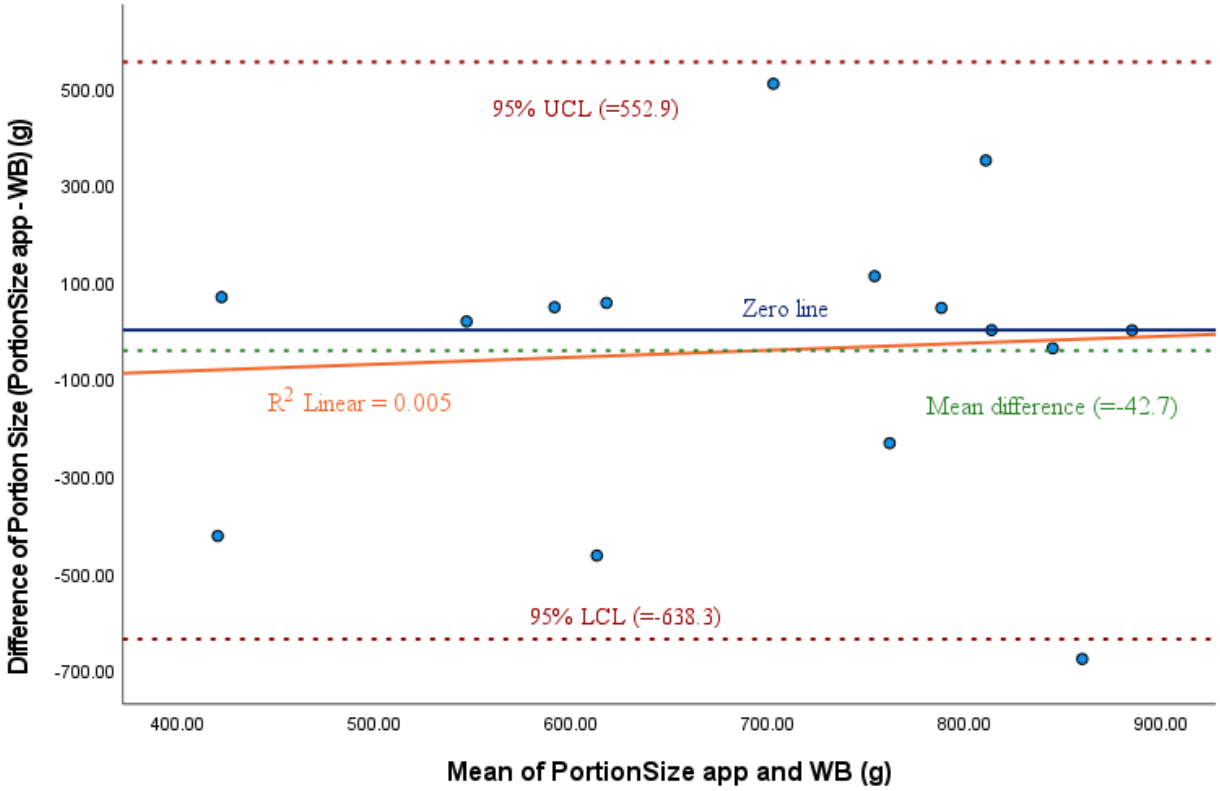
Figure 2. Bland-Altman analysis for comparing energy intake (kcal) between PortionSize and the WB method (15 meals). LCL: lower confidence limit; UCL: upper confidence limit; WB: weigh back.



Estimation of Portion Size

The mean gram weight of meals estimated with the PortionSize app was 674.3 (SD 222.8) g, and the mean estimated from the weighed meals was 716.9 (SD 207.2) g. The mean difference in the estimated gram weights of food items between the two methods was −42.7 (95% CI −638.3 to 552.9) g/meal, and both means were significantly equivalent (P=.03; Table 2 and Table S2 in Multimedia Appendix 1). The mean percent error for the estimation of gram weight was −6%. The Bland-Altman limit of agreement plot for the gram weights of food items (Figure 3) indicates a slightly positive trend (not significant) for the difference in estimated gram weights (PortionSize – weigh back) with regard to the means of both methods ([PortionSize + weigh back]/2). The results show nonsignificant bias (R^2=.005; adjusted R^2=−0.071; P=.80).

Figure 3. Bland-Altman analysis for comparing consumed food in grams (portion size) between PortionSize and the WB method (15 meals). LCL: lower confidence limit; UCL: upper confidence limit; WB: weigh back.



Estimation of Food Group Servings

The mean PortionSize-estimated total fruit servings was 0.2 (SD 0.3; Table 2). The mean weigh back method–estimated servings of fruits was 0.3 (SD 0.4), and the means were equivalent (*P*=.01). PortionSize-estimated total dairy servings (mean 0.4, SD 0.6 servings) were equivalent to the weigh back–estimated servings (mean 0.5, SD 0.6 servings; *P*=.047). Among the five food groups, the estimations of total vegetable servings had the lowest mean percent error (0%), with those for grain servings having the highest (41.7%). The results of dependent *t* tests for comparing the intake of energy, nutrient, and food group servings between the two methods—the PortionSize app and weigh back—are presented in Table S2 in Multimedia Appendix 1. Bland-Altman analyses of food groups are presented in Figures S4-S8 in Multimedia Appendix 1.

Estimation of Macronutrients and Specific Nutrients

The estimates for the mean intake of protein (*P*=.25), fat (*P*=.32), and carbohydrates (*P*=.10) were not equivalent between PortionSize and the weigh back method (Table 2). The mean differences in protein, fat, and carbohydrate intake estimates between PortionSize and the weigh back method were 2.6 (SD 17.4) g, 5.4 (SD 18.7) g, and 6.9 (SD 33.8) g, respectively. We found significant equivalence in estimations of total sugar (*P*=.049) and potassium (*P*=.049) intake between PortionSize and the weigh back method. Among the macronutrient estimations, total fat estimations had the highest mean percent error (19.9%).

User Satisfaction and CSUQ Survey

Table 3 shows that of the 15 participants, 12 (80%) were satisfied or extremely satisfied with the PortionSize app, and 11 (73%) were similarly satisfied with the easiness of the PortionSize app for recording portion sizes. Moreover, 13 (87%) participants marked “very much” for how much the iPhone training helped them to prepare for using the PortionSize app.

The CSUQ survey indicated that 11 (73%) participants strongly agreed that they could become productive quickly by using the app, and the information provided for the app was easy to understand (Table S3 in Multimedia Appendix 1). The mean score from the CSUQ was 35.3 (SD 13.5).

Table 3. Participants' satisfaction with the PortionSize app (N=15).

Questions	Score, n (%)					
	1[a]	2	3	4	5	6[b]
1. How satisfied are you with the PortionSize app for recording portion sizes?	N/A[c]	N/A	N/A	3 (20)	8 (53)	4 (27)
2. How satisfied are you with the *Portion Summary* tab of the PortionSize app?	N/A	N/A	2 (13)	1 (7)	6 (40)	6 (40)
3. How satisfied are you with the *Take Photo* tab of the PortionSize app?	N/A	N/A	N/A	4 (27)	6 (40)	5 (33)
4. How satisfied are you with the *Forgot Meal* tab of the PortionSize app?	N/A	1 (7)	N/A	N/A	4 (27)	10 (67)
5. How satisfied are you with the *Settings* tab of the PortionSize app?	N/A	N/A	N/A	1 (7)	4 (27)	10 (67)
6. How easy was it to use the PortionSize app for recording portion sizes?	N/A	N/A	N/A	4 (27)	7 (47)	4 (27)
7. How easy was it to capture images and record portion sizes?	N/A	N/A	1 (7)	7 (47)	4 (27)	3 (20)
8. How easy was it to use the *Forgot Meal* tab to describe portions?	N/A	N/A	2 (13)	1 (7)	4 (27)	8 (53)
9. How much did the iPhone training help prepare you for using the PortionSize app?	N/A	N/A	N/A	N/A	2 (13)	13 (87)
10. How appropriate were the PortionSize templates superimposed on your food items?	N/A	N/A	N/A	1 (7)	8 (53)	6 (40)

[a]Scores of 1 indicated "extremely dissatisfied," "very difficult," "not at all," and "not appropriate" for questions 1 to 5, questions 6 to 8, question 9, and question 10, respectively.

[b]Scores of 6 indicated "extremely satisfied," "extremely easy," "very much," and "very appropriate" for questions 1 to 5, questions 6 to 8, question 9, and question 10, respectively.

[c]N/A: not applicable.

Discussion

Principal Findings

In this small pilot and feasibility study, we collected preliminary validation data for the PortionSize app. The results indicate that the estimations of energy intake and the intake of energy-contributing nutrients (carbohydrates, protein, and fat) from the PortionSize app were not equivalent to those estimated via weigh back. The mean percent error of the energy intakes estimated by the PortionSize app and from weighed food was 12.7% and fell within the ranges of 8% to 30% for 24-hour dietary recall and 1.3% to 47% for diet histories, food records, and food frequency questionnaires [1]. PortionSize's error for estimating gram weight intake was smaller (−6%). There are mixed results that suggest that app-based assessment methods either underestimate or overestimate energy intake when compared with the doubly labeled water method and traditional methods, such as dietary records [1,9]. Traditional methods, such as self-reported 24-hour dietary recall, have significantly underreported energy intake when compared with 7-day food weigh records [34]. A systematic review and meta-analysis study found that image-based dietary assessments underestimated energy intake by 20% (range 0%-37%) when compared with the doubly labeled water method; however, the study showed no significant difference in energy intakes estimated via traditional methods (such as 24-hour dietary recall) and the RFPM [35]. We observed in this pilot study that the CI for the mean difference in energy intake estimations between the two methods—the PortionSize app and weigh back—crossed 0 (Table S2 in Multimedia Appendix 1).

The results of this pilot study indicate that the estimated portion sizes (g) of food from the PortionSize app was equivalent to the portion sizes that were estimated via weigh back, and the mean percent error between the two methods was −6%. Inaccurate portion size estimates necessarily result in inaccurate food intake estimates, and approximately 50% of the error in self-reported food intake is due to inaccurate portion size estimates, with missing data likely accounting for the majority of the remaining error [36]. Furthermore, consumers' difficulties with estimating portion sizes are a barrier to correctly measuring energy intake [37]. The PortionSize app integrates visual templates that enable users or consumers to accurately quantify portion sizes in real time [7]. However, users need to be discreet when selecting the right FNDDS food codes within the PortionSize app to match with the food items in order to correctly estimate their food intake.

Equivalent estimations of total fruit and total dairy servings were found between PortionSize and the weigh back method. We cross-checked the outlier values for the vegetable, grain, and protein group servings. A participant experienced app glitches when using the PortionSize app and thus could not correctly record plate waste for the grain group servings (outlier value is reported in Figure S6 in Multimedia Appendix 1). In addition, 1 participant mistakenly reported percent plate waste when using the PortionSize app and thus generated an outlier value for the protein group servings (Figure S8 in Multimedia Appendix 1). The estimations of total vegetable servings were not equivalent; however, the mean percent error was 0%. This reflects that there is no fundamental issue with the PortionSize app; however, app improvements are needed for the correct estimation of food intake.

The USDA provides recommendations in terms of food groups, such as fruits, vegetables, grains, protein, and dairy [20]. To our knowledge, PortionSize app is the first food intake assessment tool that provides immediate feedback on energy intake and food group servings. Such feedback can help users track whether they are meeting the recommended daily intake of energy, fruits, vegetables, grains, dairy, and protein.

Additionally, obtaining real-time feedback on food selection will provide an opportunity for users to modify their food intake and thus improve their food intake behavior [3,38,39]. Unspecific or delayed feedback is not as effective as real-time feedback at inducing behavior change, since behavior change is promoted by receiving immediate and specific feedback based on objective data that are temporally associated with a target behavior [38]. Food intake and dietary patterns outline an individual's nutrition status because food intake encompasses energy intake; nutrient intake (macronutrients and micronutrients, including vitamins and minerals); and the consumption of different food groups, such as fruits and vegetables [1,17,40].

A survey-based study that focused on the perceived burdens of and preferences for traditional methods, the RFPM, and PortionSize indicated that 67.3% of participants preferred to use the RFPM, 51.9% preferred the PortionSize app, 48% preferred food records, and 32.9% preferred 24-hour dietary recall. Nevertheless, a significantly higher percentage of older adults (aged ≥65 years) preferred using food records and 24-dietary recall when compared to other participants (aged <65 years) [7]. Older adults often perceive barriers and difficulties in using mobile health apps [41], and this could be one of the reasons that older adults typically prefer traditional methods. Participants in the survey-based study perceived the RFPM to less burdensome compared to the PortionSize app [7]; however, the RFPM needs trained human raters to analyze food images and assess food intake [16]. On the other hand, the PortionSize app has more embedded features for measuring food intake, including those for capturing images of food selection and plate waste, identifying foods, and estimating portion size, and these may challenge users or consumers when using the PortionSize app. Low health and nutrition literacy could be potential barriers to accurately estimating portion size [40]. The PortionSize app provides real-time feedback on food intake and food group servings; therefore, we expect that the advantages of using PortionSize app will promote users' willingness to use the PortionSize app and accept the challenges. We also expect that future validation studies will support this hypothesis.

Limitations

The purpose of this pilot and feasibility study was to examine the initial validity of the PortionSize app and inform power for future validation studies of the PortionSize app. This study has several limitations that need to be acknowledged. First, it had a small but appropriate sample size for a pilot study. Second, most of the participants were female (11/15, 73%) and highly educated (college or above: 14/15, 93%). Third, there was limited representation from different racial or ethnic groups. Fourth, while all BMI categories were targeted, the effects across different BMI categories could not be examined due to the small sample size. Fifth, each participant estimated a single meal with limited food items; however, they did not estimate food intake over a long duration (ie, days or weeks). Sixth, the PortionSize app was also designed to estimate and provide feedback on alcohol consumption; however, we did not analyze alcohol consumption in this study. Lastly, the PortionSize app was recently developed, and the team continues to debug the app and improve its functionality. This likely impacted user satisfaction. Future studies with a large sample are needed to examine differences in food intake estimations from the PortionSize app between men and women, among BMI categories, and among different ethnic groups.

Conclusions

The findings from this pilot study suggest that the PortionSize app has promise for estimating food intake in real time. With some improvements, it is hoped that the PortionSize app will become sufficiently accurate, so that it can be used by participants to modify their food intake in real time (ie, when they are selecting foods) and how much of each food they eat during a meal or snack.

Acknowledgments

The authors would like to thank the participants for their time and effort in this study. We are grateful to Randy Ullrich for his assistance in app development.

This work was supported by the National Institute of Diabetes and Digestive and Kidney Diseases (R01, grant DK124558; P30, grant DK072476; T32, grant DK064584) and the Louisiana Clinical and Translational Science Center (U54, grant GM104940).

Authors' Contributions

SS, CPL, CKM, and JWA formulated the research questions. SB, CKM, and JWA designed this study. CKM and JWA carried out this study. SS, CPL, and SB analyzed the data. SS, CPL, SB, CKM, and JWA interpreted the findings. SS, CPL, SB, CKM, and JWA wrote this paper.

Conflicts of Interest

The intellectual property related to the PortionSize app is owned by the Louisiana State University System and Pennington Biomedical Research Center. Authors CKM and JWA are the inventors of the technology and are employed by the Louisiana State University System and Pennington Biomedical Research Center.

Multimedia Appendix 1
Supplementary material.

[DOCX File , 374 KB - formative_v6i6e38283_app1.docx]

References

1. Burrows TL, Ho YY, Rollo ME, Collins CE. Validity of dietary assessment methods when compared to the method of doubly labeled water: A systematic review in adults. Front Endocrinol (Lausanne) 2019 Dec 17;10:850 [FREE Full text] [doi: 10.3389/fendo.2019.00850] [Medline: 31920966]
2. Ambrosini GL, Hurworth M, Giglia R, Trapp G, Strauss P. Feasibility of a commercial smartphone application for dietary assessment in epidemiological research and comparison with 24-h dietary recalls. Nutr J 2018 Jan 09;17(1):5 [FREE Full text] [doi: 10.1186/s12937-018-0315-4] [Medline: 29316930]
3. Martin CK, Correa JB, Han H, Allen HR, Rood JC, Champagne CM, et al. Validity of the Remote Food Photography Method (RFPM) for estimating energy and nutrient intake in near real-time. Obesity (Silver Spring) 2012 Apr;20(4):891-899 [FREE Full text] [doi: 10.1038/oby.2011.344] [Medline: 22134199]
4. McClung HL, Ptomey LT, Shook RP, Aggarwal A, Gorczyca AM, Sazonov ES, et al. Dietary intake and physical activity assessment: Current tools, techniques, and technologies for use in adult populations. Am J Prev Med 2018 Oct;55(4):e93-e104 [FREE Full text] [doi: 10.1016/j.amepre.2018.06.011] [Medline: 30241622]
5. Schembre SM, Liao Y, O'Connor SG, Hingle MD, Shen SE, Hamoy KG, et al. Mobile ecological momentary diet assessment methods for behavioral research: Systematic review. JMIR Mhealth Uhealth 2018 Nov 20;6(11):e11170 [FREE Full text] [doi: 10.2196/11170] [Medline: 30459148]
6. Teixeira V, Voci SM, Mendes-Netto RS, da Silva DG. The relative validity of a food record using the smartphone application MyFitnessPal. Nutr Diet 2018 Apr;75(2):219-225. [doi: 10.1111/1747-0080.12401] [Medline: 29280547]
7. Höchsmann C, Fearnbach N, Dorling JL, Fazzino TL, Myers CA, Apolzan JW, et al. Preference, expected burden, and willingness to use digital and traditional methods to assess food and alcohol intake. Nutrients 2021 Sep 24;13(10):3340 [FREE Full text] [doi: 10.3390/nu13103340] [Medline: 34684341]
8. Kouvari M, Mamalaki E, Bathrellou E, Poulimeneas D, Yannakoulia M, Panagiotakos DB. The validity of technology-based dietary assessment methods in childhood and adolescence: a systematic review. Crit Rev Food Sci Nutr 2021;61(7):1065-1080. [doi: 10.1080/10408398.2020.1753166] [Medline: 32308010]
9. Shinozaki N, Murakami K. Evaluation of the ability of diet-tracking mobile applications to estimate energy and nutrient intake in Japan. Nutrients 2020 Oct 29;12(11):3327 [FREE Full text] [doi: 10.3390/nu12113327] [Medline: 33138088]
10. Zhao X, Xu X, Li X, He X, Yang Y, Zhu S. Emerging trends of technology-based dietary assessment: a perspective study. Eur J Clin Nutr 2021 Apr;75(4):582-587. [doi: 10.1038/s41430-020-00779-0] [Medline: 33082535]
11. Shim JS, Oh K, Kim HC. Dietary assessment methods in epidemiologic studies. Epidemiol Health 2014 Jul 22;36:e2014009 [FREE Full text] [doi: 10.4178/epih/e2014009] [Medline: 25078382]
12. Naska A, Lagiou A, Lagiou P. Dietary assessment methods in epidemiological research: current state of the art and future prospects. F1000Res 2017 Jun 16;6:926 [FREE Full text] [doi: 10.12688/f1000research.10703.1] [Medline: 28690835]
13. Martin CK, Nicklas T, Gunturk B, Correa JB, Allen HR, Champagne C. Measuring food intake with digital photography. J Hum Nutr Diet 2014 Jan;27 Suppl 1(0 1):72-81 [FREE Full text] [doi: 10.1111/jhn.12014] [Medline: 23848588]
14. McConahy KL, Smiciklas-Wright H, Birch LL, Mitchell DC, Picciano MF. Food portions are positively related to energy intake and body weight in early childhood. J Pediatr 2002 Mar;140(3):340-347. [doi: 10.1067/mpd.2002.122467] [Medline: 11953733]
15. Olafsdottir AS, Hörnell A, Hedelin M, Waling M, Gunnarsdottir I, Olsson C. Development and validation of a photographic method to use for dietary assessment in school settings. PLoS One 2016 Oct 06;11(10):e0163970. [doi: 10.1371/journal.pone.0163970] [Medline: 27711120]
16. Martin CK, Han H, Coulon SM, Allen HR, Champagne CM, Anton SD. A novel method to remotely measure food intake of free-living individuals in real time: the remote food photography method. Br J Nutr 2009 Feb;101(3):446-456 [FREE Full text] [doi: 10.1017/S0007114508027438] [Medline: 18616837]
17. Young LR, Nestle M. Expanding portion sizes in the US marketplace: implications for nutrition counseling. J Am Diet Assoc 2003 Feb;103(2):231-234. [doi: 10.1053/jada.2003.50027] [Medline: 12589331]
18. Almiron-Roig E, Solis-Trapala I, Dodd J, Jebb SA. Estimating food portions. Influence of unit number, meal type and energy density. Appetite 2013 Dec;71:95-103 [FREE Full text] [doi: 10.1016/j.appet.2013.07.012] [Medline: 23932948]
19. Champagne CM, Bray GA, Kurtz AA, Monteiro JBR, Tucker E, Volaufova J, et al. Energy intake and energy expenditure: a controlled study comparing dietitians and non-dietitians. J Am Diet Assoc 2002 Oct;102(10):1428-1432. [doi: 10.1016/s0002-8223(02)90316-0] [Medline: 12396160]
20. What is MyPlate? MyPlate U.S. Department of Agriculture. URL: https://www.myplate.gov/eat-healthy/what-is-myplate [accessed 2022-01-12]
21. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
22. World Medical Association. World Medical Association Declaration of Helsinki. Ethical principles for medical research involving human subjects. Bull World Health Organ 2001;79(4):373-374 [FREE Full text] [Medline: 11357217]

23. Hertzog MA. Considerations in determining sample size for pilot studies. Res Nurs Health 2008 Apr;31(2):180-191. [doi: 10.1002/nur.20247] [Medline: 18183564]
24. Plucker A, Thomas DM, Broskey N, Martin CK, Schoeller D, Shook R, et al. Adult energy requirements predicted from doubly labeled water. Int J Obes (Lond) 2018 Aug;42(8):1515-1523. [doi: 10.1038/s41366-018-0168-0] [Medline: 30026590]
25. Roe BE, Apolzan JW, Qi D, Allen HR, Martin CK. Plate waste of adults in the United States measured in free-living conditions. PLoS One 2018 Feb 14;13(2):e0191813. [doi: 10.1371/journal.pone.0191813] [Medline: 29444094]
26. Food and nutrient database for dietary studies (2017-2018). U.S. Department of Agriculture. URL: https://www.ars.usda.gov/northeast-area/beltsville-md-bhnrc/beltsville-human-nutrition-research-center/food-surveys-research-group/docs/fndds-download-databases/ [accessed 2022-01-13]
27. Lewis JR. IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. Int J Hum Comput Interact 1995;7(1):57-78. [doi: 10.1080/10447319509526110]
28. Lewis JR. Usability: Lessons learned … and yet to be learned. Int J Hum Comput Interact 2014 Jul 10;30(9):663-684. [doi: 10.1080/10447318.2014.930311]
29. Lewis JR. Psychometric evaluation of the PSSUQ using data from five years of usability studies. Int J Hum Comput Interact 2002;14(3-4):463-488. [doi: 10.1080/10447318.2002.9669130]
30. Al Ayubi SU, Parmanto B, Branch R, Ding D. A persuasive and social mHealth application for physical activity: A usability and feasibility study. JMIR Mhealth Uhealth 2014 May 22;2(2):e25 [FREE Full text] [doi: 10.2196/mhealth.2902] [Medline: 25099928]
31. English LL, Dunsmuir D, Kumbakumba E, Ansermino JM, Larson CP, Lester R, et al. The PAediatric Risk Assessment (PARA) mobile app to reduce postdischarge child mortality: Design, usability, and feasibility for health care workers in Uganda. JMIR Mhealth Uhealth 2016 Feb 15;4(1):e16 [FREE Full text] [doi: 10.2196/mhealth.5167] [Medline: 26879041]
32. Walker E, Nowacki AS. Understanding equivalence and noninferiority testing. J Gen Intern Med 2011 Feb;26(2):192-196 [FREE Full text] [doi: 10.1007/s11606-010-1513-8] [Medline: 20857339]
33. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986 Feb 08;1(8476):307-310. [Medline: 2868172]
34. Jayawardena R. Comparison dietary assessment methods in Sri Lankan adults: use of 24-hour dietary recall and 7-day weighed intake. BMC Nutr 2016 Mar 21;2(18):1-4 [FREE Full text] [doi: 10.1186/s40795-016-0059-5]
35. Ho DKN, Tseng SH, Wu MC, Shih CK, Atika AP, Chen YC, et al. Validity of image-based dietary assessment methods: A systematic review and meta-analysis. Clin Nutr 2020 Oct;39(10):2945-2959. [doi: 10.1016/j.clnu.2020.08.002] [Medline: 32839035]
36. Beasley J, Riley WT, Jean-Mary J. Accuracy of a PDA-based dietary assessment program. Nutrition 2005 Jun;21(6):672-677. [doi: 10.1016/j.nut.2004.11.006] [Medline: 15925290]
37. Livingstone MBE, Pourshahidi LK. Portion size and obesity. Adv Nutr 2014 Nov 14;5(6):829-834 [FREE Full text] [doi: 10.3945/an.114.007104] [Medline: 25398749]
38. Schultz W. Behavioral theories and the neurophysiology of reward. Annu Rev Psychol 2006;57:87-115. [doi: 10.1146/annurev.psych.56.091103.070229] [Medline: 16318590]
39. de Roos B. Personalised nutrition: ready for practice? Proc Nutr Soc 2013 Feb;72(1):48-52. [doi: 10.1017/S0029665112002844] [Medline: 23199675]
40. Chaudhry BM, Schaefbauer C, Jelen B, Siek KA, Connelly K. Evaluation of a food portion size estimation interface for a varying literacy population. 2016 May Presented at: CHI'16: CHI Conference on Human Factors in Computing Systems; May 7-12, 2016; San Jose, California p. 5645-5657. [doi: 10.1145/2858036.2858554]
41. Askari M, Klaver NS, van Gestel TJ, van de Klundert J. Intention to use medical apps among older adults in the Netherlands: Cross-sectional study. J Med Internet Res 2020 Sep 04;22(9):e18080 [FREE Full text] [doi: 10.2196/18080] [Medline: 32624465]

Abbreviations

CSUQ: Computer System Usability Questionnaire
FNDDS: Food and Nutrition Dataset for Dietary Studies
IRB: institutional review board
PBRC: Pennington Biomedical Research Center
RFPM: remote food photography method
USDA: United States Department of Agriculture

Edited by A Mavragani; submitted 28.03.22; peer-reviewed by M Allman-Farinelli, C Jacob; comments to author 08.05.22; revised version received 24.05.22; accepted 26.05.22; published 15.06.22.

Please cite as:
Saha S, Lozano CP, Broyles S, Martin CK, Apolzan JW
Assessing the Initial Validity of the PortionSize App to Estimate Dietary Intake Among Adults: Pilot and Feasibility App Validation Study
JMIR Form Res 2022;6(6):e38283
URL: https://formative.jmir.org/2022/6/e38283
doi:10.2196/38283
PMID:35704355

Original Paper

Use of Smartphone Health Apps Among Patients Aged 18 to 69 Years in Primary Care: Population-Based Cross-sectional Survey

Sabrina Paradis[1,2], PhD; Jeremy Roussel[1], PhD; Jean-Luc Bosson[2,3], PhD; Jean-Baptiste Kern[1], PhD

[1]Department of General Practice, University Grenoble Alpes, Grenoble, France

[2]Translational Innovation in Medicine and Complexity, National Institute of Health and Medical Research, La Tronche, France

[3]Department of Public Health, University Grenoble Alpes, Grenoble, France

Corresponding Author:
Sabrina Paradis, PhD
Translational Innovation in Medicine and Complexity
National Institute of Health and Medical Research
Laboratoire TIMC, Bâtiment Taillefer
Rond-Point de la Croix de Vie, Domaine de la Merci
La Tronche, 38706
France
Phone: 33 678863377
Email: sabrina.paradis@univ-grenoble-alpes.fr

Abstract

Background: The World Health Organization has defined mobile health (mHealth) as the "use of mobile and wireless technologies to support the achievement of health objectives." Smartphones currently represent one of the main media forms for mHealth democratization. Health apps can be an interesting tool for changing health behaviors. However, their use in France is still poorly documented.

Objective: The main aim of this study was to evaluate the frequency of use of health apps among patients consulting in the primary care setting in France. The secondary aims were to evaluate the use of health apps according to the sociodemographic and medical characteristics of patients and to determine their use.

Methods: A population-based cross-sectional survey was carried out between November 2017 and January 2018 in the Grenoble area of France among patients aged between 18 and 69 years who were consulting at 13 primary care physician offices. Patients were provided with anonymous paper self-questionnaires. The main criterion for participation was the use of a smartphone health app, defined for the purpose of this study as any app supporting patients in efforts to be healthy.

Results: The participation rate was 49.27% (739/1500; 95% CI 46.7%-51.8%). The smartphone use was estimated at 82.6% (597/723; 95% CI 79.6%-85.2%). Of 597 smartphone owners, 47.7% (283/595; CI 43.6%-51.6%) used at least one smartphone health app. Health apps identified in this study were mainly related to wellness, prevention, and fitness (66.1%), as well as medication, treatments, and follow-up care (50.0%). The main factors associated with health app use were: use of social networks (odds ratio [OR] 3.4, 95% CI 2.1-5.3), age under 30 years (OR 2.7, CI 1.4-4.9), city size between 5001 and 10,000 inhabitants (OR 1.8, CI 1.1-2.8), and city size more than 10,000 inhabitants (OR 2.1, CI 1.4-3.2).

Conclusions: In this survey, nearly one out of two patients reported the use of smartphone health apps, which are currently focused on wellness, prevention, and fitness, and are largely used by the younger population.

Trial Registration: ClinicalTrials.gov NCT03351491; https://clinicaltrials.gov/ct2/show/NCT03351491

(JMIR Form Res 2022;6(6):e34882) doi:10.2196/34882

KEYWORDS

smartphone; health applications; mHealth; apps; mobile health; digital health; well-being; epidemiology; primary care; population survey; fitness; physical activity; health behavior; patient

Introduction

Mobile health (mHealth) is generally defined as medical and public health practice supported by mobile devices such as mobile phones, patient monitoring devices, personal digital assistants, and other wireless devices. The World Health Organization has defined mHealth as the "use of mobile and wireless technologies to support the achievement of health objectives" [1].

A mobile phone is a wireless portable device that allows users to make and receive calls. Modern mobile phones are more commonly called "smartphones" because of the many digital mobile services they offer. Owing to their advanced computing and their connectivity using cellular network architecture, smartphones currently represent one of the main media forms for mHealth democratization and spread. Between 2011 and 2020 in France, smartphone ownership among people older than 12 years increased from 17% to 84%, with 94% of the population older than 12 years owning a mobile phone in 2020 [2]. Similar statistics were reported the same year in the United States, with a rate of smartphone ownership estimated at 85% [3]. This equipment evolution has been accompanied by an increase in the number and use of mobile apps. These software programs running on devices such as smartphones are preinstalled or downloadable on apps markets (eg, Google Play Store, Apple App Store). Some of these focus on health, fitness, or medical care.

In the last few decades, smartphones have radically modified our daily lives. In the field of health, patients have more access to health knowledge, with greater opportunities to improve the involvement of patients through patient-professional partnerships. Several studies have demonstrated the effectiveness of mHealth interventions to improve health behavior [4,5] in contexts such as diet adhesion, smoking cessation, increasing physical activity, and chronic disease management (eg, diabetes or hypertension). This evolution has impacted the relationship between general practitioners (GPs) and their patients [6]. Recent research suggests that patients perceive mHealth apps as useful complementary tools for self-monitoring and self-management of their health, albeit with some limits [7]. A qualitative study involving French GPs highlighted an ambivalent discourse around the prescription of mHealth apps or patients' use of apps [8]. In France, health apps must be considered as "medical devices" and must be evaluated by the National Committee for the Evaluation of Medical Devices and Health Technologies (CNEDIMTS) to obtain reimbursement [9]. Currently, to our knowledge, only one app is reimbursed by the health insurance and can be prescribed (Moovcare).

In 2015, a survey conducted in the United States showed that 58.2% of mobile phone users had a health app compared to only 19% in 2012 [10]. The proportion of smartphone owners using health apps was estimated at 20.5% in 2015 in Germany [11] and at 24.1% in 2016 in Hong Kong [12]. However, there is a paucity of data in the scientific literature concerning the use of health apps among smartphone users in France.

The main objective of this study was to evaluate the frequency of use of at least one mobile health app on a smartphone since its acquisition by patients recruited from the primary care setting in the Grenoble region of France. The secondary objectives were to collect the types of health apps used, and to analyze factors associated with the use of these apps according to sociodemographic, geographical, and medical characteristics of the studied population.

Methods

Design

This study was a population-based cross-sectional survey carried out among patients aged 18-69 years consulting GPs in their office or in a primary care center in the Grenoble area of France. Data were collected between November 2017 and January 2018 by anonymous self-administered paper questionnaires.

Sample

Inclusion criteria of the considered sample were: all adult outpatients aged between 18 and 69 years, consulting a GP and not being deprived of liberty by judicial or administrative decision or subject to a legal protection measure. All patients were eligible regardless of the reason for the consultation. The sample size was calculated using the following assumptions: precision=5%, α risk=5%, estimated percentage of smartphone users=65%, estimated health apps use in the population of smartphone users=50%. The calculated number was then increased by 60% to take nonresponses into account. The minimum required sample size was estimated at 1476 questionnaires. This number was then rounded up to 1500.

Questionnaire

A specific questionnaire was developed for the purpose of the study and distributed to participants. The questionnaire consists of 23 questions and was estimated to be completed in approximately 5 minutes. It was submitted to expert opinion (one GP, one epidemiologist, and one biostatistician) and then tested on 15 patients. Their advice and suggestions were taken into account, and the questionnaire was modified to be as understandable and relevant as possible. As health apps may be more frequently used by internet users, we opted for a paper questionnaire to avoid potential selection bias. After receiving clear oral information on the terms and objectives of the study, written information was presented on the first page. A patient's refusal to participate in the study could be indicated on the first page. A total of 1500 self-administered anonymous paper questionnaires were distributed to 35 voluntary GPs in 12 offices and in 1 primary care center. These were selected to obtain a representative sample of patients in terms of sociodemographic factors (see Multimedia Appendix 1). Convenience sampling was applied in the centers. Questionnaires were administered by the GPs or their secretaries after having checked that the inclusion criteria were respected. Once completed, they were stocked in a box in each investigation center. All boxes were collected during the last week of January 2018.

The first question collected the details of mobile phone and smartphone equipment owned. The primary outcome measure was the use of at least one mobile health app on a smartphone. In this study, a health app refers to an app supporting patients in efforts to be healthy without distinguishing between wellness

and prevention apps. Online health and well-being websites available with a smartphone were also considered to be in scope. The download and the use of smartphone apps were also measured without distinguishing between common apps and health apps. The secondary outcome measures were the digital characteristics (determined by the digital equipment and the use of social networks), sociodemographic characteristics (age, gender, socioprofessional category), geographical characteristics (determined by the postcode), and the medical characteristics (medication and long-term disease). The socioprofessional categories were derived from the nomenclature of the French National Institute for Statistical and Economic Studies (Institut National de la Statistique et des Études Économiques [INSEE]) and they were gathered in four categories (two for the labor force and two for the nonworking population) by median annual income. The types of used health apps were derived from the French National Market Research Agency (Institut Français d'Opinion Publique [IFOP]) and simplified to be more comprehensible. The following variables regarding the type of health app use were collected: treatment and follow-up care (4 classes); emergency management (2 classes); communication (3 classes); well-being, prevention, and fitness (5 classes); fertility and pregnancy (2 classes); and diagnostic assistance (2 classes) (see Table 1).

Table 1. Types of health apps.

Category	Subclass
Treatment and follow-up care	Drug information, disease information, medical parameters management (eg, blood pressure, weight), help receiving treatments
Emergency management	Warning system (eg, prerecorded emergency numbers), first-aid help (eg, basic emergency life-saving skills)
Communication	Health professional search (eg, phone books), data sharing (eg, medical mailbox), exchange about health themes
Well-being, prevention, and fitness	Smoking cessation, fitness, nutrition and weight loss, stress management, sleeping help
Fertility and pregnancy	Ovulation schedule, pregnancy calendar
Diagnostic assistance	Symptom information, self-diagnosis help

Data Analysis

The database input was performed by two authors, and entries were verified for 10% of the questionnaires. Statistical analyses were performed with Stata 15.0. Quantitative variables are expressed as mean (SD) and were compared with a *t*-test after normality was confirmed or with the Mann-Whitney *U* test if normality was not confirmed. Qualitative variables are expressed in numbers with percentages and were compared with the χ^2 test (or Fisher exact test in the event of small numbers). Statistical testing was performed with an α risk equal to .05. Multivariate analysis was performed by logistic regression with a subgroup analysis. Variables selected were those with $P<.20$ in univariate analysis. Odds ratios (ORs) were calculated with the 95% CIs for each variable. Missing data were incorporated in the statistical analysis.

Ethics Approval and Registration

This study was approved by the Committee for the Protection of Persons (Comité de Protection des Personnes) Sud-Ouest et Outre-Mer III (2017-A01647-46) and by the Correspondent for Protection of Personal Data (Correspondant Informatique et Libertés) of University Grenoble Alps (0987763). The protocol was registered at ClinicalTrials.gov (NCT03351491).

Results

The response rate of the distributed questionnaires was estimated at 49.3% (739/1500; 95% CI 46.7%-51.8%). Eight patients refused to give their consent (Table 2). Among our study participants, the proportion of mobile phone owners was estimated at 96.6% (714/731; 95% CI 96.3%-98.5%). Smartphone use was estimated at 82.6% (597/723; 95% CI 79.6%-85.2%). A total of 134 patients were excluded from the final analysis: 119 did not use a smartphone, 7 did not know what kind of mobile phone they used, and 8 questionnaires contained missing data concerning the use of a smartphone. A total of 597 questionnaires were then included in the statistical analysis.

The sample characteristics are summarized in Table 3. The average age of the patients was 41.1 (SD 5.6) years and 65% of the sample were women (95% CI 61.1%-68.7%).

Most of the participants (455/594, 76.6%; 95% CI 73.0%-79.8%) used apps on their smartphone, and 47.7% (283/595; 95% CI 43.6%-51.6%) used at least one health app since acquiring their smartphone. All of these users (283/283, 100.0%) had already downloaded an app (health app or not) on apps markets. The most commonly used app types were well-being, prevention, or fitness apps (185/280, 66.1%; 95% CI 60.3%-71.4%); apps about drugs, treatment, and follow-up care (140/280, 50.0%; 95% CI 44.2%-56.0%); apps for communication with health professionals (98/280, 35.0%; 95% CI 29.6%-40.8%); and apps about fertility and pregnancy (58/280, 20.7%; 95% CI 16.3%-25.9%).

The least used apps were those for emergency management (27/280, 9.6%; 95% CI 6.9%-13.7%) and apps for diagnosis help and symptoms data (26/280, 9.3%; 95% CI 6.4%-13.3%). The use of more than one app was reported by 46.6% (132/280; 95% CI 40.9%-52.5%) of patients. Women used more health apps about fertility and pregnancy; well-being, prevention, and fitness; and drugs, treatment, and follow-up care than men (Table 4).

The main origins of awareness of the existence of the used apps were the social circle (114/265, 43%; 95% CI 37.1%-49.1%) and media (internet, TV, newspapers, and other media channels) (103/265, 38.9%; 95% CI 33.1%-44.9%). In addition, 8.7% (23/265; 95% CI 5.8%-12.7%) of GPs recommended health apps to patients, whereas only 6.0% (16/265; 95% CI 3.7%-9.6%) and 3.7% (10/265; 95% CI 2.0%-6.9%) of other physicians and other health professionals, respectively, recommended these apps. Other sources were given for 19.2% (51/265; 95% CI 14.9%-24.5%) of the respondents, including preinstalled health apps (10/39) or downloaded apps following personal research (11/39). The frequency of use of these apps was indicated to be rare for 31.1% (84/270; 95% CI 25.8%-36.9%), monthly for 21.5% (58/270; 95% CI 16.9%-26.8%), weekly for 16.3% (44/270; 95% CI 12.3%-21.2%), and daily for 31.1% (84/270; 95% CI 25.8%-36.9%) of respondents.

Univariate analysis revealed that the use of health apps was associated with a young population, mainly female, living in larger cities, and the use of social network(s) (all $P<.001$). A potential association between health app use and socioprofessional category ($P<.001$) was also identified.

By contrast, the presence of chronic conditions and the number of treatments were not associated with the use of health apps (Table 5).

The logistic regression model displayed an association between the use of health apps and the use of social networks; age under 30 years; being a woman; living in a city with 5001-10,000 inhabitants; living in a city with more than 10,000 inhabitants; and occupying an executive position, intellectual profession, or having an intermediate occupation (Table 6).

Table 2. Participation by center.

Center	Distributed (N=1500)	Collected (n=739), n (%)	Participants (% total sample)	Women (n=473), n (%)	Refusal to participate, n (%)	Smartphone user (n=597), n (%)
1	43	13 (30.2)	1.76	7 (1.5)	0 (0)	10 (1.7)
2	172	43 (25.0)	5.82	26 (5.5)	2 (25)	31 (5.2)
3	129	62 (48.1)	8.39	40 (8.5)	0 (0)	52 (8.7)
4	172	127 (73.8)	17.19	87 (18.4)	0 (0)	107 (17.9)
5	43	33 (76.7)	4.47	27 (5.7)	0 (0)	30 (5.0)
6	43	38 (88.4)	5.14	26 (5.5)	1 (12.5)	36 (6.0)
7	167	96 (57.5)	12.99	59 (12.5)	3 (37.5)	80 (13.4)
8	129	122 (94.6)	16.51	80 (16.9)	0 (0)	94 (15.7)
9	172	94 (54.7)	12.72	47 (9.9)	0 (0)	68 (11.4)
10	86	45 (52.3)	6.09	28 (5.9)	1 (12.5)	35 (5.9)
11	43	29 (67.4)	3.92	21 (4.4)	1 (12.5)	24 (4.0)
12	43	19 (44.2)	2.57	13 (2.7)	0 (0)	16 (2.7)
13	258	18 (7.0)	2.44	12 (2.5)	0 (0)	14 (2.3)

Table 3. Characteristics of the sample.

Characteristic	Value
Age (years) (n=596), n (%)	
<30	143 (24.0)
30-39	140 (23.5)
40-49	134 (22.5)
50-69	179 (30.0)
Women (n=597), n (%)	388 (65.0)
Use of social network(s) (n=597), n (%)	464 (72.7)
Chronic disease, n (%)	
Overall (n=592)	204 (34.5)
Cancer (n=575)	18 (3.1)
Diabetes (n=577)	28 (4.8)
Psychiatric diseases (n=577)	19 (3.3)
Cardiac diseases (n=577)	36 (6.7)
Rheumatologic diseases (n=577)	29 (5.0)
Pulmonary diseases (n=577)	14 (2.4)
Renal diseases (n=577)	5 (0.9)
Other chronic diseases (n=577)	80 (13.9)
Medications (n=583) (number per day), mean (SD)	0.9 (0.7)
Population (n=597), n (%)	
0-1000	7 (1.2)
1001-5000	190 (31.8)
5001-10,000	177 (29.6)
>10,000	223 (37.3)
Socioprofessional category (n=595), n (%)	
Farmers, craftspeople, storekeepers, managers, workers, or employees	295 (49.6)
Executive, intermediate, or intellectual professions	159 (26.7)
Retired	52 (8.7)
Student or unemployed	89 (15.0)

Table 4. Use and type of health apps used by gender.

Use of health app	Respondents, n	Women, n (%)	Men, n (%)	*P* value
Overall users	283	206 (73.0)	77 (27.0)	<.001
Drugs, treatment, and follow-up care	140	102 (72.9)	38 (27.1)	<.001
Emergency management	27	17 (63.0)	10 (37.0)	.84
Communication with health professionals	98	69 (70.4)	29 (29.6)	.25
Well-being, prevention, and fitness	185	134 (72.4)	51 (27.6)	<.001
Fertility and pregnancy	58	57 (98.3)	1 (1.7)	<.001
Diagnosis help and symptoms data	26	21 (80.8)	5 (19.2)	.10

Table 5. Univariate analysis of characteristics associated with the use of health apps.

Characteristic	Health apps use (n=283)	No health apps use (n=312)	*P* value
Age (years) (n=596), mean (SD)	37.7 (7.8)	44.2 (7.7)	<.001
Women (n=388), n (%)	206 (72.8)	180 (57.7)	<.001
Use of social network(s) (n=594), n (%)	244 (86.2)	190 (60.9)	<.001
Chronic disease , n (%)			
Overall (n=592)	90 (32.0)	114 (36.9)	.12
Cancer (n=575)	10 (3.6)	8 (2.7)	.35
Diabetes (n=577)	10 (3.6)	18 (6.0)	.12
Psychiatric diseases (n=577)	10 (3.6)	9 (3.0)	.44
Cardiac diseases (n=577)	14 (5.1)	22 (7.4)	.16
Rheumatologic diseases (n=577)	16 (5.8)	13 (4.4)	.28
Pulmonary diseases (n=577)	7 (2.5)	7 (2.4)	.55
Renal diseases (n=577)	3 (1.1)	2 (0.7)	.47
Other chronic diseases (n=577)	39 (14.1)	41 (13.8)	.50
Number of treatments (n=583), mean (SD)	0.9 (0.1)	1.0 (0.1)	.23
Population (n=596), n (%)			<.001
0-1000	1 (0.3)	6 (1.9)	
1001-5000	69 (24.4)	121 (38.8)	
5001-10,000	89 (31.4)	87 (27.8)	
>10,000	124 (43.8)	98 (31.4)	
Socioprofessional category (n=595), n (%)			.001
Farmers, craftspeople, storekeepers, managers, workers or employees	141 (50.0)	152 (48.9)	
Executives, intermediate or intellectual professions	83 (29.4)	76 (24.4)	
Retired	13 (4.6)	39 (12.5)	
Students or unemployed	45 (16.0)	44 (14.1)	

Table 6. Factors associated with the use of health apps (logistic regression model).

Characteristic	Adjusted OR[a]	95% CI
Age (years)		
18-29 (n=143)	2.68	1.45-4.94
30-39 (n=139)	1.22	0.70-2.15
40-49 (n=134)	1.46	0.85-2.51
50-69 (n=178)	Reference	Reference
Socioprofessional category		
Farmers, craftspeople, storekeepers, managers, workers, or employees	Reference	Reference
Executives, intermediate or intellectual professions	1.71	1.10-2.65
Retired	0.80	0.36-1.80
Students or unemployed	0.76	0.45-1.29
Women	1.77	1.21-2.59
Chronic disease	1.28	0.83-1.96
Population		
0-5,000	Reference	Reference
5001-10,000	1.81	1.15-2.84
>10,000	2.10	1.37-3.22
Social network(s) use	3.36	2.12-5.34

[a]OR: odds ratio.

Discussion

Principal Findings

In this survey, 82.6% of participants owned a smartphone, 76.6% used apps, and 46.7% used at least one health app. These users of health apps tended to be under 30 years old, women, use social networks, belong to a higher socioprofessional category, and live in larger cities compared with other smartphone owners. Over 84% of the population over 12 years old in France was reported to use a smartphone in 2020 (representing an increase from 77% in 2019) [13], with 85% of US adults estimated to use a smartphone in 2020 (representing an increase from 81% in 2019) [14]. The smartphone user proportion of 82.6% (597/723; 95% CI 79.6%-85.2%) found in this study was thus slightly higher than the estimate of French data from 2019. This difference could be explained by the characteristics of our sample with a young, tech-friendly population. Indeed, we did not include people over 69 years old. As this older population is less equipped in smartphones than people between 18 and 69 years, not including them should have caused a higher proportion of smartphone owners in our sample.

No estimation of the use of health apps in France was found in the scientific literature. Therefore, the results of this survey had to be compared to those of other countries. To our knowledge, the more recently published results about the rate of use of general health apps concern the United States and Germany, with both studies conducted in 2015, along with one study conducted in Hong Kong in 2016 [12], providing estimates of health apps use among smartphone owners of 20.5% [15], 58.2% [10], and 24.1% [12], respectively. These differences could be attributed to the different sample recruitment strategies used. Our sample was recruited from GPs in their office or in a primary care center, whereas the previous studies were conducted through population-based surveys. Thus, selection bias may be possible, because health consciousness might be more important among patients than among the general population [16].

This disparity could be also attributed to differences in sample characteristics, and especially the mean age. Indeed, our studied sample had a mean age of 41 years, which is close to that of the US survey (40 years); however, the mean age of the sample in Germany was 57 years and 16.1% of the participants of the Hong Kong survey were elderly patients (≥65 years). Our results suggested that the younger population had a greater tendency to use health apps than older participants. In addition, a recent systematic review regarding factors influencing use of mHealth apps showed that mHealth apps are mainly used by young people [17]. Thus, by excluding patients over 69 years of age in our survey, the mean age decreased while the overall use of health apps increased in comparison with those reported for the general population.

In our results, the most frequently used health apps were those concerning well-being, prevention, and fitness (185/280, 66.1%), and those related to treatment, drugs, and follow-up care (140/280, 50.0%). These could not be directly compared to the previous studies of Ernsting et al [11] and Krebs and Duncan [10], as they measured the reasons for downloading apps but did not directly compare the types of apps downloaded. In the German survey [11], the participants mainly reported using apps to support changes in smoking cessation (44.5%), healthy diet

(38.6%), weight loss (23.2%), and physical activity (17.08%), whereas the participants of the US survey [10] reported physical activity tracking (52.8%), nutrition tracking (47.6%), and the desire to lose weight (46.8%) as the main reasons to download health apps. The results of the Hong Kong survey [12] also showed that physical activity tracking (67%), logging health records (43%), and tracking health measures (30.2%) were the most frequent reasons for downloading and using apps. Despite the differences of measurement methods, these data seem to reinforce our results and confirm a greater interest for prevention, fitness, and well-being apps. This same conclusion was reached by the Pew Research Center in 2012 in that health apps related to fitness (38%), diet (31%), and weight management (12%) were more common, whereas medication management apps were only used by 2% of respondents in a US survey [18].

Concerning the frequency of health apps use, this study found either rare (84/270, 31.1%) or daily use (84/270, 31.1%) as the most common answers. These opposite results might be explained by the gradual loss of interest for the app(s) over time; patients likely use their apps more often initially due to the appeal of novelty, which then decreases progressively. This hypothesis was also supported by the US survey [10], in which 45.7% of health app users no longer used the health app, including 40% who indicated that this was due to a loss of interest.

Our results demonstrated an association of being aged under 30 years with the use of mHealth apps. The same conclusions were obtained in other surveys. This association could be explained by the high smartphone ownership rate among younger patients and their tendency to download more apps; 80% among those 18-24 years of age and 72% among those 25-39 years of age used their mobile phones to download app(s) in France compared with only 44% among those 40-59 years of age and 20% among those 60-69 years of age [9].

Social networks use was also associated with the use of health apps on a smartphone (OR 3.36, 95% CI 2.1-5.3). We can suppose that this is because social network users are more inclined to download apps in general, and thus more health apps, than nonusers.

We also found an association of health app use with gender. Indeed, women appeared to use health apps more than men. This might be due to the use of fertility and pregnancy apps. This association could be qualified because this link was not demonstrated in previously reported surveys, and because women seemed to be overrepresented in our sample (65%, 95% CI 61.1%- 68.7%). However, the previous surveys were conducted in the general population, whereas ours focused on people consulting a GP. Women tend to consult GPs more often than men, especially before 55 years of age, and therefore the characteristics of our sample could explain this difference. Moreover, the same association was reported in the Pew Research Center survey [18].

Our results also suggested that people who occupied an executive position, an intellectual profession, or an intermediate occupation had a tendency to use health apps more than others. This could be partly explained by the higher use of smartphones among people in this socioprofessional category and by their greater use of apps in general. An association was also found between health apps use, a high income, and higher level of education in previous surveys [13,15]. This therefore strengthens our conclusions, as the identified socioprofessional category comprises those earning the most money and having followed the highest level of education.

Moreover, living in cities with more than 5000 inhabitants was also associated with greater use of health apps. Indeed, the multivariate analysis demonstrated that the bigger the city, the stronger the association. Thus, cities with more than 10,000 inhabitants were more strongly associated with apps use (OR 2.10, 95% CI 1.37-3.22) than those with 5001-10,000 inhabitants (OR 1.81, 95% CI 1.15-2.84). The size of cities was not explored in other surveys. These data might be explained by the concentration of students in Grenoble (8.0% vs 2.0% in the same region in 2010), who are younger and thus more likely to use apps, and of higher-income people such as executives or those with intermediate occupations (14.0% and 16.6%, respectively, vs 11.9% and 16% in the whole region) in these areas.

In contrast to the findings of Ernsting et al [11], we did not find an association of health apps use with having a chronic disease. This result could be explained by our inclusion criteria. Indeed, patients over 69 years old, who represent the population most affected by chronic diseases, were not included in our sample. Moreover, we did not consider BMI or ethnicity in our study, which were associated with health apps use in the other surveys. A survey in the general population or without limitation of age would be interesting to further study the associations of these factors.

Limitations

Despite the agreement of these results with previous data from the literature, some limitations should be taken into consideration with respect to interpreting the results of this study. First, this study was not randomized and the participation was voluntary, which could be a limitation of sample representativeness. Furthermore, we did not explore the use of health apps in the population over 60 years old. This choice of inclusion criterion might have created a bias in assessing the associations with chronic disease and the number of treatments, which were found in some previous studies. Finally, this was a self-response survey, which can cause misunderstandings and missing data for some questions in the questionnaire, despite performing pilot tests with several patients before official recruitment. Considering these limitations, these results from a regional survey are consequently hard to extend to the general population.

This study was performed prior to the COVID-19 pandemic. During the pandemic period, eHealth, mHealth, and telemedicine were more widely adopted for crisis management and as a preventive measure to increase clinical care [19,20]. Thus, a similar study should be performed to determine the impact of the pandemic on the use of eHealth. An additional element in this survey was that there were no questions related to the use of health apps by GPs; thus, it would be interesting to evaluate

the potential link between this practice and the patients' use of mHealth.

In spite of these limitations, precautions were taken to ensure the relevance of the results. The sample was composed of patients from 13 centers that represented various types of primary care practices. The substantial sample size further instilled confidence in the robustness of the statistical methods used, with sufficient power to detect statistically significant differences. Moreover, we used paper questionnaires rather than online surveys. This decision was taken so as to avoid selection bias of people that use the internet regularly, who would also be more likely to own a smartphone and to use health apps. These precautions were used to ensure the strength of our findings, and our results indeed are in agreement with studies performed in other countries.

Conclusion

This work confirmed the wide extent of smartphone use among 82.6% of the sample. Moreover, an important use of health apps was identified in our sample, with nearly one out of two smartphone users reporting downloading and using health apps. The users of these apps had a tendency to be younger than 30 years old, to be women, to live in bigger cities, to use more social networks, and to be in a high socioprofessional category. Currently, the use of mHealth is mainly limited to well-being and prevention apps, including fitness and weight loss apps. However, health apps have demonstrated their potential in changing health behavior and could thus become a new tool for health professionals to help improve their patients' health conditions. Physicians could design health apps or advise patients about them, aiming to improve their relationship and facilitating the development of partnership in their care. Nevertheless, health apps could be more efficient if they reached a broader population. Making these apps easier to use could also help to democratize them. Accordingly, health apps could play a major role in health care, especially for people affected by chronic conditions or for elderly people. App developers are therefore encouraged to improve their software to make their apps more accessible to a maximum of patients. However, the development of health apps and the huge and increasing number of app marketplaces come with a problem of reduced quality; indeed, important discrepancies can be observed between apps, and patients do not currently have the appropriate information and formal evidence of their effectiveness. This major challenge is currently under discussion in France and worldwide [21].

Acknowledgments

This study was supported by the Department of General Practice and Translational Innovation in Medicine and Complexity Laboratory.

Authors' Contributions

SP and JR conceived the study. JBK directed the study. JLB reviewed the study design. SP revised the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Details of health centers participating in the study.
[DOCX File , 28 KB - formative_v6i6e34882_app1.docx]

References

1. Ryu S. Book Review: mHealth: New Horizons for Health through Mobile Technologies: Based on the Findings of the Second Global Survey on eHealth (Global Observatory for eHealth Series, Volume 3). Healthc Inform Res 2012;18(3):231. [doi: 10.4258/hir.2012.18.3.231]
2. Barometre du numerique edition 2021. Enquête sur la diffusion des technologies de l'information et de la communication dans la société française. ARCEP. URL: https://www.arcep.fr/uploads/tx_gspublication/rapport-barometre-numerique-edition-2021.pdf [accessed 2022-02-17]
3. Perrin A. Mobile technology and home broadband. Pew Research Center. 2021 Jun 03. URL: https://www.pewresearch.org/internet/2021/06/03/mobile-technology-and-home-broadband-2021/ [accessed 2022-02-17]
4. Han M, Lee E. Effectiveness of mobile health application use to improve health behavior changes: a systematic review of randomized controlled trials. Healthc Inform Res 2018 Jul;24(3):207-226 [FREE Full text] [doi: 10.4258/hir.2018.24.3.207] [Medline: 30109154]
5. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
6. Müssener U. Digital encounters: human interactions in mHealth behavior change interventions. Digit Health 2021;7:20552076211029776 [FREE Full text] [doi: 10.1177/20552076211029776] [Medline: 34262783]

7. Vo V, Auroy L, Sarradon-Eck A. Patients' perceptions of mHealth apps: meta-ethnographic review of qualitative studies. JMIR Mhealth Uhealth 2019 Jul 10;7(7):e13817 [FREE Full text] [doi: 10.2196/13817] [Medline: 31293246]
8. Sarradon-Eck A, Bouchez T, Auroy L, Schuers M, Darmon D. Attitudes of general practitioners toward prescription of mobile health apps: qualitative study. JMIR Mhealth Uhealth 2021 Mar 04;9(3):e21795 [FREE Full text] [doi: 10.2196/21795] [Medline: 33661123]
9. Hassanaly P, Dufour JC. Analysis of the regulatory, legal, and medical conditions for the prescription of mobile health applications in the United States, the European Union, and France. Med Devices (Auckl) 2021;14:389-409. [doi: 10.2147/MDER.S328996] [Medline: 34853541]
10. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. JMIR Mhealth Uhealth 2015 Nov 04;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
11. Ernsting C, Stühmann LM, Dombrowski SU, Voigt-Antons J, Kuhlmey A, Gellert P. Associations of Health App Use and Perceived Effectiveness in people with cardiovascular diseases and diabetes: population-based survey. JMIR Mhealth Uhealth 2019 Mar 28;7(3):e12179 [FREE Full text] [doi: 10.2196/12179] [Medline: 30920383]
12. Shen C, Wang MP, Chu JT, Wan A, Viswanath K, Chan SSC, et al. Health app possession among smartphone or tablet owners in Hong Kong: population-based survey. JMIR Mhealth Uhealth 2017 Jun 05;5(6):e77 [FREE Full text] [doi: 10.2196/mhealth.7628] [Medline: 28583905]
13. Baromètre du numérique 2019. Enquête sur la diffusion des technologies de l'information et de la communication dans la société française en 2019. ARCEP. URL: https://www.arcep.fr/uploads/tx_gspublication/rapport-barometre-num-2019.pdf [accessed 2020-02-05]
14. Anderson M. Mobile technology and home broadband 2019. Pew Research Center. 2019. URL: https://www.pewinternet.org/2019/06/13/mobile-technology-and-home-broadband-2019/ [accessed 2019-10-25]
15. Ernsting C, Dombrowski SU, Oedekoven M, O Sullivan JL, Kanzler M, Kuhlmey A, et al. Using smartphones and health apps to change and manage health behaviors: a population-based survey. J Med Internet Res 2017 Apr 05;19(4):e101 [FREE Full text] [doi: 10.2196/jmir.6838] [Medline: 28381394]
16. Jayanti RK, Burns AC. The antecedents of preventive health care behavior: an empirical study. J Acad Market Sci 1998 Jan 01;26(1):6-15. [doi: 10.1177/0092070398261002]
17. Wang C, Qi H. Influencing factors of acceptance and use behavior of mobile health application users: systematic review. Healthcare (Basel) 2021 Mar 22;9(3):357 [FREE Full text] [doi: 10.3390/healthcare9030357] [Medline: 33809828]
18. Fox S, Duggan M. Mobile health 2012. Pew Research Center. 2012 Nov 08. URL: https://www.pewresearch.org/internet/2012/11/08/mobile-health-2012/ [accessed 2021-02-05]
19. Petrazzuoli F, Kurpas D, Vinker S, Sarkisova V, Eleftheriou A, Żakowicz A, et al. COVID-19 pandemic and the great impulse to telemedicine: the basis of the WONCA Europe Statement on Telemedicine at the WHO Europe 70th Regional Meeting September 2020. Prim Health Care Res Dev 2021 Dec 13;22:e80 [FREE Full text] [doi: 10.1017/S1463423621000633] [Medline: 34895388]
20. Bitar H, Alismail S. The role of eHealth, telehealth, and telemedicine for chronic disease patients during COVID-19 pandemic: A rapid systematic review. Digit Health 2021;7:20552076211009396 [FREE Full text] [doi: 10.1177/20552076211009396] [Medline: 33959378]
21. Mathews SC, McShea MJ, Hanley CL, Ravitz A, Labrique AB, Cohen AB. Digital health: a path to validation. NPJ Digit Med 2019 May 13;2(1):38. [doi: 10.1038/s41746-019-0111-3] [Medline: 31304384]

Abbreviations

CNEDIMTS: National Committee for the Evaluation of Medical Devices and Health Technologies
GP: general practitioner
IFOP: Institut Français d'Opinion Publique (French National Market Research Agency)
INSEE: Institut National de la Statistique et des Études Économiques (French National Institute for Statistical and Economic Studies)
mHealth: mobile health
OR: odds ratio

Edited by A Mavragani; submitted 11.11.21; peer-reviewed by S Rostam Niakan Kalhori, F Velayati, C Baxter; comments to author 22.12.21; revised version received 09.03.22; accepted 04.04.22; published 16.06.22.

Please cite as:
Paradis S, Roussel J, Bosson JL, Kern JB
Use of Smartphone Health Apps Among Patients Aged 18 to 69 Years in Primary Care: Population-Based Cross-sectional Survey
JMIR Form Res 2022;6(6):e34882
URL: https://formative.jmir.org/2022/6/e34882
doi:10.2196/34882
PMID:35708744

Original Paper

Tailoring Mobile Data Collection for Intervention Research in a Challenging Context: Development and Implementation in the Malakit Study

Yann Lambert[1], MD, MEng; Muriel Galindo[1], MSc; Martha Sua´rez-Mutis[2], MD; Louise Mutricy[1], ADN; Alice Sanna[1], MD; Laure Garancher[3], MEng; Hedley Cairo[4], MD; Helene Hiwat[4], PhD; Jane Bordalo Miller[5], BA; Jose´ Hermenegildo Gomes[5], BA; Paola Marchesini[6], MD; Antoine Adenis[1], MD, PhD; Mathieu Nacher[1], MD, PhD; Stephen Vreden[7], MD, PhD; Maylis Douine[1], MD, PhD

[1]Centre d'Investigation Clinique Antilles-Guyane, Institut national de la santé et de la recherche médicale (Inserm 1424), Centre Hospitalier de Cayenne Andre´e Rosemon, Cayenne, French Guiana

[2]Parasitic Disease Laboratory, Institute Oswaldo Cruz, Foundation Oswaldo Cruz, Rio de Janeiro, Brazil

[3]The Ink Link, Paris, France

[4]Malaria Program, Ministry of Health of Suriname, Paramaribo, Suriname

[5]Desenvolvimento, Prevenção, Acompanhamento e Cooperação de Fronteiras, Oiapoque, Brazil

[6]Malaria Technical Group, Vector Transmissible and Zoonotic Diseases Coordination, Ministry of Health, Brasilia, Brazil

[7]Foundation for Scientific Research Suriname, Paramaribo, Suriname

Corresponding Author:
Yann Lambert, MD, MEng
Centre d'Investigation Clinique Antilles-Guyane
Institut national de la santé et de la recherche médicale (Inserm 1424)
Centre Hospitalier de Cayenne Andre´e Rosemon
Avenue des Flamboyants
Cayenne, 97300
French Guiana
Phone: 594 594 39 48 64
Email: yann.lambert@ch-cayenne.fr

Abstract

Background: An interventional study named Malakit was implemented between April 2018 and March 2020 to address malaria in gold mining areas in French Guiana, in collaboration with Suriname and Brazil. This innovative intervention relied on the distribution of kits for self-diagnosis and self-treatment to gold miners after training by health mediators, referred to in the project as facilitators.

Objective: This paper aims to describe the process by which the information system was designed, developed, and implemented to achieve the monitoring and evaluation of the Malakit intervention.

Methods: The intervention was implemented in challenging conditions at five cross-border distribution sites, which imposed strong logistical constraints for the design of the information system: isolation in the Amazon rainforest, tropical climate, and lack of reliable electricity supply and internet connection. Additional constraints originated from the interaction of the multicultural players involved in the study. The Malakit information system was developed as a patchwork of existing open-source software, commercial services, and tools developed in-house. Facilitators collected data from participants using Android tablets with ODK (Open Data Kit) Collect. A custom R package and a dashboard web app were developed to retrieve, decrypt, aggregate, monitor, and clean data according to feedback from facilitators and supervision visits on the field.

Results: Between April 2018 and March 2020, nine facilitators generated a total of 4863 form records, corresponding to an average of 202 records per month. Facilitators' feedback was essential for adapting and improving mobile data collection and monitoring. Few technical issues were reported. The median duration of data capture was 5 (IQR 3-7) minutes, suggesting that electronic data capture was not taking more time from participants, and it decreased over the course of the study as facilitators become more experienced. The quality of data collected by facilitators was satisfactory, with only 3.03% (147/4849) of form records requiring correction.

Conclusions: The development of the information system for the Malakit project was a source of innovation that mirrored the inventiveness of the intervention itself. Our experience confirms that even in a challenging environment, it is possible to produce good-quality data and evaluate a complex health intervention by carefully adapting tools to field constraints and health mediators' experience.

Trial Registration: ClinicalTrials.gov NCT03695770; https://clinicaltrials.gov/ct2/show/NCT03695770

(JMIR Form Res 2022;6(6):e29856) doi:10.2196/29856

KEYWORDS

malaria; Guiana Shield; information system; mobile data collection; Open Data Kit; ODK

Introduction

Plasmodium falciparum and *Plasmodium vivax* malaria remain endemic in the Region of the Guiana Shield [1]. In French Guiana, the main reservoir of the parasites is the population of 10,000 to 15,000 gold miners working clandestinely in the Amazon forest [2]. Mostly originating from Brazil, they have virtually no access to health care while working on the mining sites [3]. In case of malaria-like symptoms, they frequently use under-the-counter antimalarials, without diagnosis and with poor adherence. This behavior, in combination with high mobility within the forest and across borders, may jeopardize the efforts to control malaria in the region of the Guiana Shield, paving the way for the emergence of antimalarial-resistant *Plasmodium* parasites [4,5]. To address this concern, French Guiana, Brazil, and Suriname jointly implemented an interventional study called Malakit from April 2018 to March 2020. This innovative collaborative project investigated the feasibility and effectiveness of distributing kits for self-diagnosis and self-treatment to gold miners after being trained on how to use them by health mediators [6,7].

Evaluating the effectiveness and transferability of such a novel strategy is always crucial, but the complex environment in which Malakit was implemented made this even more challenging [8]. The evaluation of the Malakit strategy was based on (1) the comparison of before-and-after cross-sectional estimates of appropriate behavior of gold miners with regard to malaria care and (2) the longitudinal monitoring of participants' adherence to correct kit use and safety. This pragmatic study design was chosen to ensure feasibility in this particular context and because the limited level of evidence produced needed support from data of the best possible quality. An information system was, thus, developed to address strong field constraints, while meeting ethical and regulatory standards, with a mobile health (mHealth) approach to data capture, management, and monitoring.

In this paper, we describe the process by which the Malakit information system was designed, developed, and implemented to achieve the monitoring and evaluation of the Malakit intervention. Through quantitative and qualitative feedback, we present our experience with mobile data collection (MDC) tools and how they may be used successfully by health mediators, hereafter referred to as facilitators.

Methods

Context and Needs of the Malakit Study

Field Constraints

French military forces are fighting illegal gold mining to limit its expansion in French Guiana. Clandestine gold miners enter the French territory through crossing points at the borders with Brazil and Suriname. These "resting sites" are strategically located in Brazil and Suriname, outside the borders of French Guiana and beyond the reach of the French military forces. They provide logistical and economical support to the gold mining system [5].

The Malakit intervention relied on facilitators who were responsible for the distribution of kits and the training of participants. These facilitators were chosen for their knowledge of the gold miners' community and their ability to communicate with them. The hiring, training, supervision, and different missions of Malakit facilitators are detailed in previous articles [7,9]. Kits were distributed at four fixed resting sites along the borders of French Guiana, with the continuous presence of two facilitators. One additional distribution site was set up in Anamoestraat, an area well frequented by Brazilian gold miners in Paramaribo, the capital city of Suriname. Furthermore, occasional mobile missions were organized at three secondary distribution sites selected for their strategic importance for the gold miners' movements to and from mining areas in French Guiana [6,7]. In Brazil and Suriname, facilitators received administrative and logistical support from one local supervisor who was in contact with the study coordination team in Cayenne. Figure 1 shows the location of the distribution sites in Suriname and Brazil, and the staff involved in the Malakit study. As described in Table 1, the geographical and logistical context varied significantly among distribution sites and impacted the work of facilitators. The Malakit information system needed to take these local constraints into account.

Figure 1. Distribution sites and staff of the Malakit study, Guiana Shield, April 2018-March 2020.

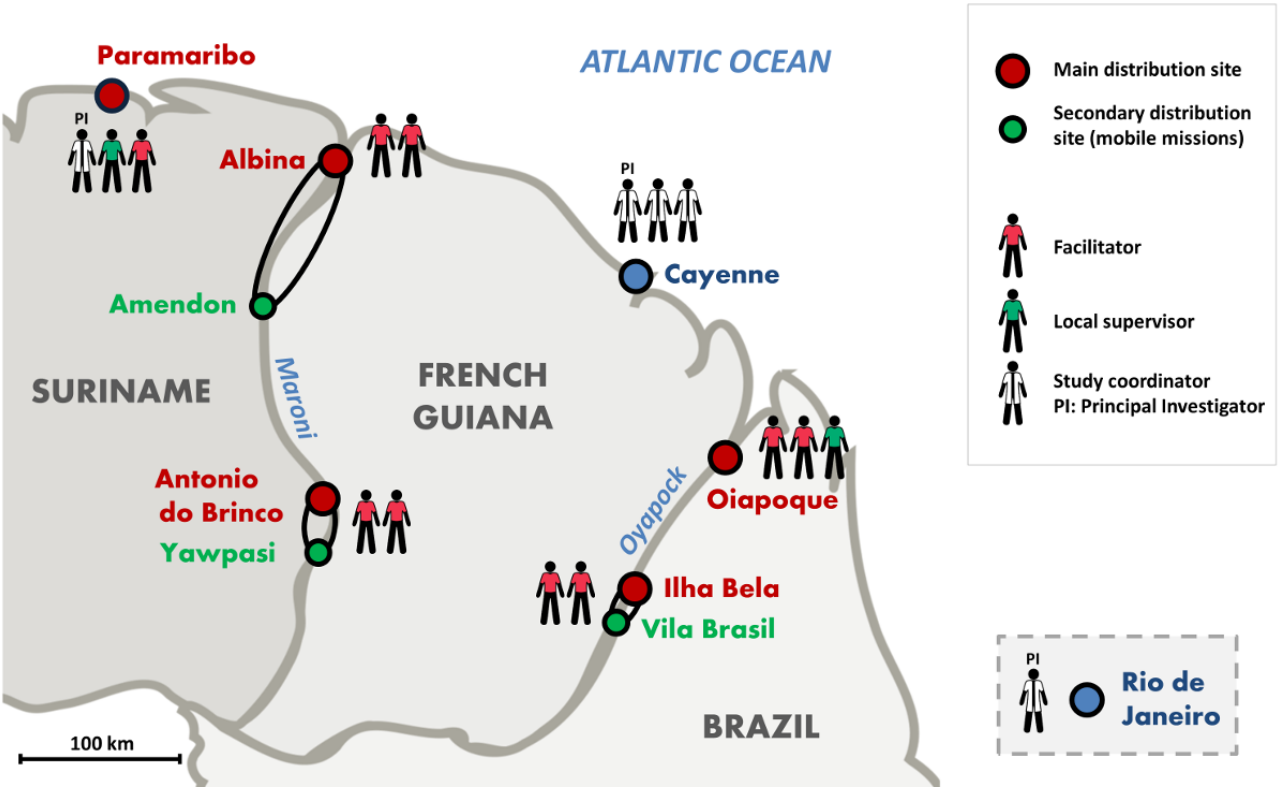


Table 1. Geographical and logistical context of distribution sites in the Malakit study, Guiana Shield, April 2018-March 2020.

Location and distribution site	Distribution mode	Environment	Access	Internet bandwidth	Electrical supply	Number of facilitators[a]
Maroni River, Surinamese border						
Albina	Fixed	Urban	Road, boat	Good	Good	2
Amendon	Mobile mission	Forest, isolated	Boat only	None	None	1 (Albina)
Antonio do Brinco	Fixed	Urban, isolated	Plane, boat	Poor	Evening only	2
Yawpasi	Mobile mission	Forest, isolated	Plane, boat	None	None	1 (Antonio do Brinco)
Suriname, capital city						
Paramaribo	Fixed	Urban	Road, plane	Excellent	Excellent	1
Oyapock River, Brazilian border						
Oiapoque	Fixed	Urban	Road, boat, plane	Good	Good, occasional power cuts	2
Ilha Bela	Fixed	Forest, very isolated	Boat only	None	Evening only	2
Vila Brasil	Mobile mission	Urban, very isolated	Boat only	Poor	Evening only	2 (Ilha Bela)

[a]The origin of the facilitators involved in mobile missions at secondary distribution sites is shown in parentheses.

Prototyping of the Information System

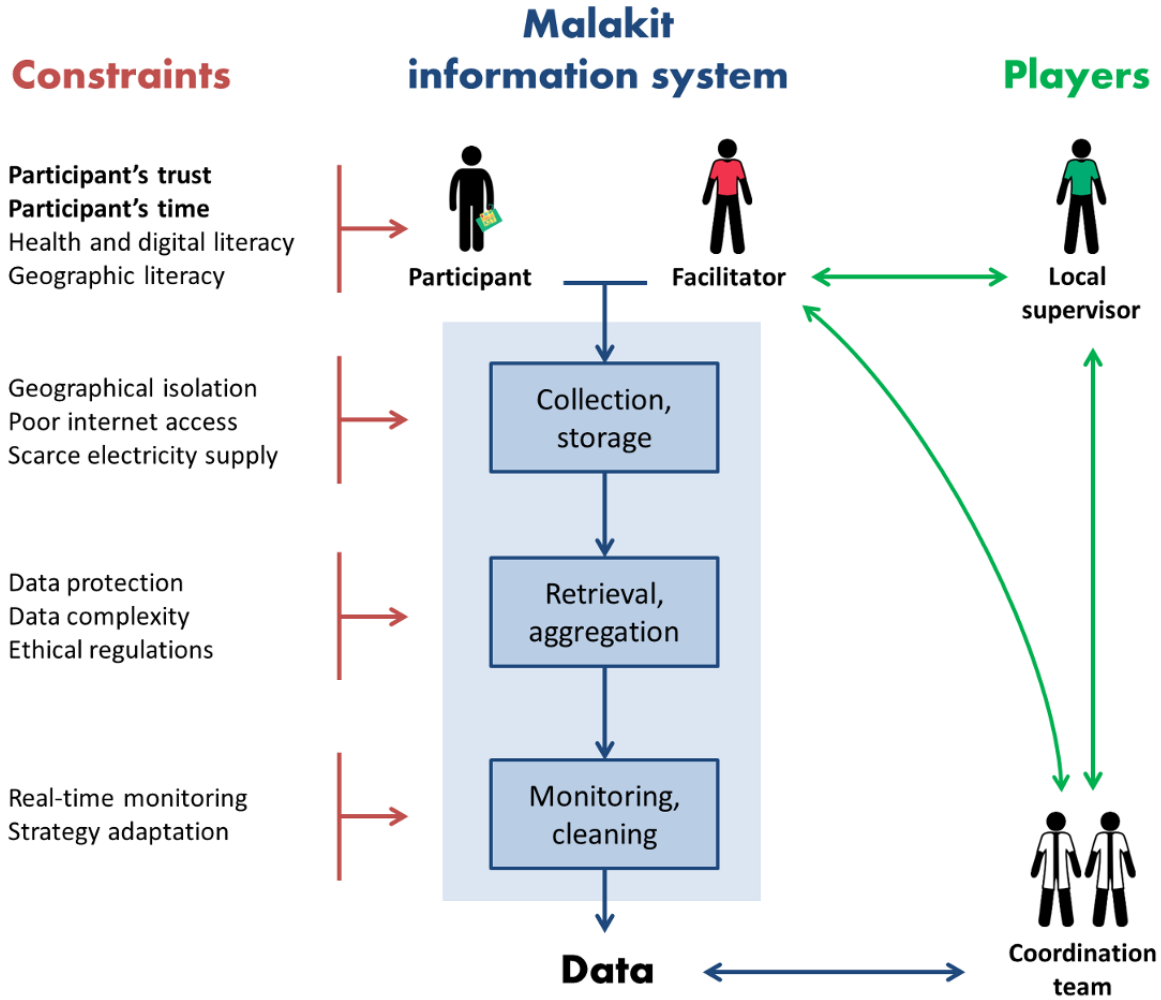
The gold mining community actively contributed to the development of Malakit. A participatory approach helped to prototype and validate the contents of the intervention (ie, kit design, training steps, and training materials, such as illustrations and videos) during multiple field missions at the resting sites in 2017 and early 2018 with groups of gold miners and key actors of the community [7,10].

The information system involved participants of the study, nine facilitators, two local supervisors, and the study coordination team. The principal investigators of the study were located in Brazil, Suriname, and French Guiana. The information system development and maintenance relied on one member of the coordination team in Cayenne, in close relationship to the study sponsor. The interaction of these different players in the Malakit project was another source of complexity added to the heterogeneity of the settings of distribution sites, as illustrated by the variety of languages used for communication (ie, Portuguese, English, French, and Dutch). Facilitators showed varying levels of health, digital, and geographic literacy, and little to no experience in participant enrollment and data

collection [7]. To ensure the quality of the data and allow for real-time monitoring, the design of the Malakit information system had to adapt to these multiple constraints. Figure 2 shows the influence of these constraints on each component of the information system and on its users.

Participants' trust and available time were regarded as the most critical constraints to take into account. Gold miners going to French Guiana wait for any opportunity to cross the border and have little time to spend at resting sites. In addition, because of their clandestine status, gold miners may be wary of any action promoted by the French authorities, and data collection for research may be confused with police intelligence. It was, therefore, very important throughout the study to be transparent about the use of the data. The hiring of facilitators from the gold mining community as an interface between the coordination team and the participants helped establish a climate of trust. Any communication about the study clearly stated its objectives and the purpose of data collection. Several components of the information system were designed to preserve this trust: participants' anonymization, questionnaire design (eg, no questions regarding future destinations and freedom to not answer), and data encryption. Before the start of the study, a first version of the information system was presented to the scientific committee of the Malakit project, which included one facilitator with several years of experience as a community health worker in the Malaria Service Deliverer network in Suriname. She helped rephrase questions, helped identify questions that could appear or would likely appear sensitive to the participants, and insisted on keeping the questionnaires as short as possible to avoid wasting time.

Figure 2. Constraints influencing the design of the Malakit information system (Malakit study, Guiana Shield, April 2018-March 2020).



Paperless Data Collection and Digital Device Selection

The choice of electronic data capture (EDC) over paper was obvious. The simultaneous collection of longitudinal data from several thousand participants, by five different teams of facilitators in locations sometimes only accessible by boat or plane, dissuaded the use of paper questionnaires. The first reason was logistical: paper-based data collection would result in additional handling and sorting, heavy weight of paper, and high risk of loss and damage during transportation by boat in the rain. A second reason was the complexity and highly variable length of questionnaires with many questions that would likely be skipped if using paper questionnaires. Use of paper would also have required extra time and staff for data entry, which would be incompatible with our need for swift access to data to monitor the safety of the study.

The device used for EDC needed to scan barcodes with a camera, and this determined the choice of tablets or smartphones. Tablets with a larger screen were preferred to allow facilitators to display videos, drawings, and a

demonstration of the Malakit app during the training of participants [7,11]. The lack of stable and reliable internet access motivated the choice of offline data collection. At distant sites, electricity was only available in the evening for 4 to 5 hours with the use of a petrol generator. Upon choosing the electronic device for data collection, battery life was, therefore, a major selection criterion: the device had to have a battery life greater than 48 hours in case of prolonged energy shortage. Hence, a tablet with a 7-inch screen was selected for its maximum battery life of 100 hours: Galaxy Tab A 7.0 (model SM-T280, 2016, Android 5.1; Samsung) [12]. The advantage of a 7-inch versus a 10-inch tablet was also a lower weight, making it more ergonomic for a prolonged use at arm's length. The selection of an Android device constrained the choice of the EDC app.

Data Collection and Storage

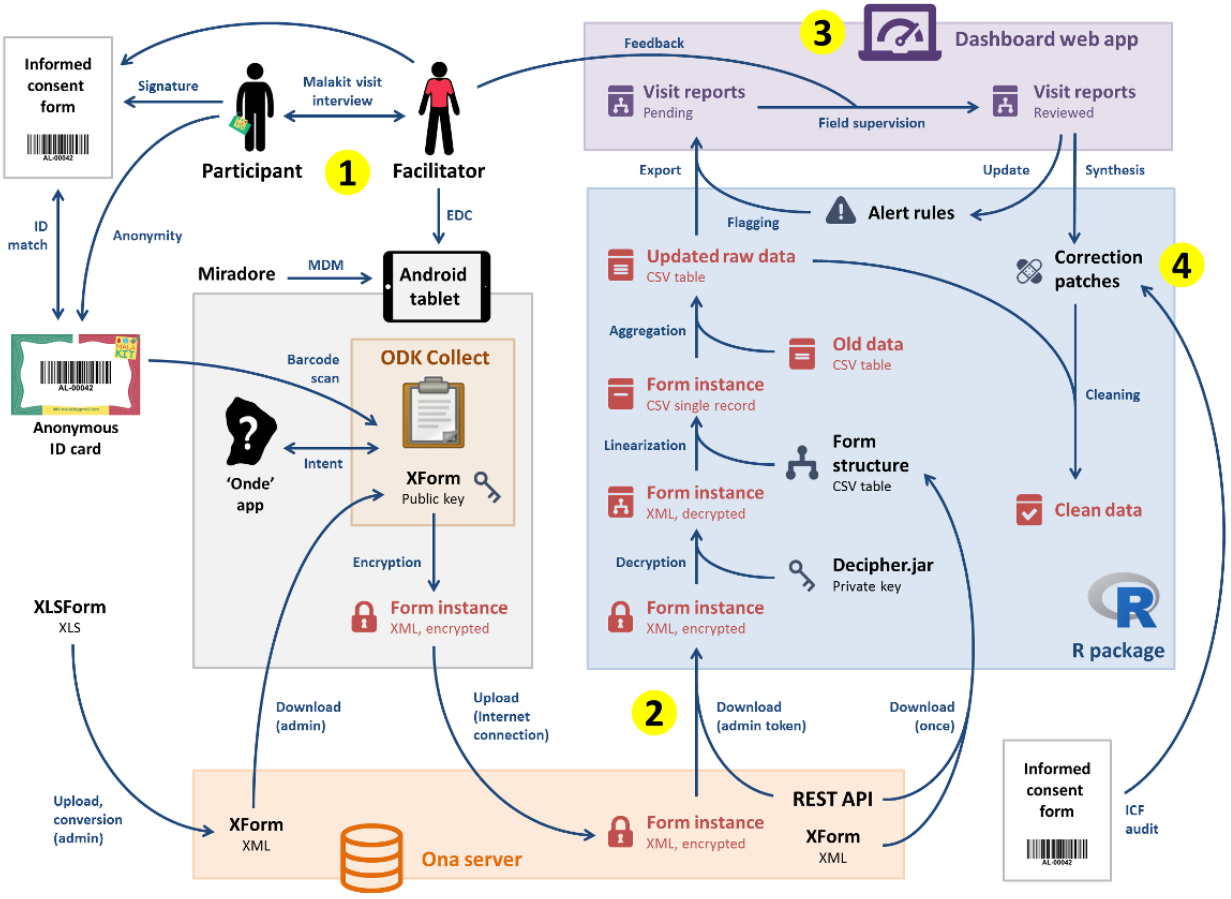
ODK (Open Data Kit) is an open-source ecosystem dedicated to data collection and management in challenging environments; it is actively developed and maintained by a worldwide community [13,14]. ODK Collect is an app that allows for offline data collection from mobile Android devices. Data may be encrypted and retrieved manually from the device or uploaded to a distant server implementing ODK Aggregate when internet access is available. ODK Collect met many requirements of the Malakit project, in particular, data encryption to comply with good clinical practice and the European Union's General Data Protection Regulation, which are rigorous processes in the management of personal health data [15].

Deploying and maintaining an instance of ODK Aggregate on the study sponsor's server was too expensive and complex for a single research project. Online services such as Ona and KoBoToolbox are alternatives to ODK Aggregate that were developed to assist humanitarian and research projects in low- and middle-income countries [16,17]. We selected Ona for its ability to manage several projects and allocate different user rights for form edition, data submission, and data access. Opting for a paid plan granted the capacity to handle sufficient numbers of weekly form submissions during the 2 years of the study; this also established a commercial bond securing data safety and server maintenance.

Development of the Malakit Information System

The workflow of the final information system developed for the Malakit study is described in Figure 3.

Figure 3. Information system of the Malakit study (collection and flow of data), Guiana Shield, April 2018-March 2020. (1) Mobile data collection by facilitators with Android tablets and storage on the Ona server. (2) Data retrieval, decryption, and aggregation with the MalakitR package. (3) Monitoring of visit reports with the Malakit dashboard web app. (4) Data cleaning with the MalakitR package. admin: administrator; EDC: electronic data capture; ICF: informed consent form; MDM: mobile device management; ODK: Open Data Kit; REST API: representational state transfer application programming interface; XForm: form standard used by ODK; XLS: Microsoft Excel spreadsheet.



Form Design for Electronic Data Capture

Malakit was designed as a longitudinal study with two different types of visits: a first inclusion visit to train the participants on the proper use of the kit, and subsequent follow-up visits to debrief participants on their episodes of symptoms of malaria and potential uses of the kit since their inclusion in the study [6,7]. Two main questionnaires were designed for the inclusion and the follow-up visits, respectively. The inclusion questionnaire covered sociodemographic data, mobility, gold mining activity, and the result of the malaria rapid diagnostic test (RDT) performed by the participants during their training. The follow-up questionnaire focused on malaria episodes experienced by participants since their last visit with a facilitator and their use of the kit. The main information collected was the severity of symptoms, the use of an RDT, the medication taken, the duration of treatment, and the experience of side effects. The follow-up questionnaire was more complex than the inclusion questionnaire because of the possibility of a participant to report up to five past episodes of symptoms of malaria, therefore requiring the repetition of related variables.

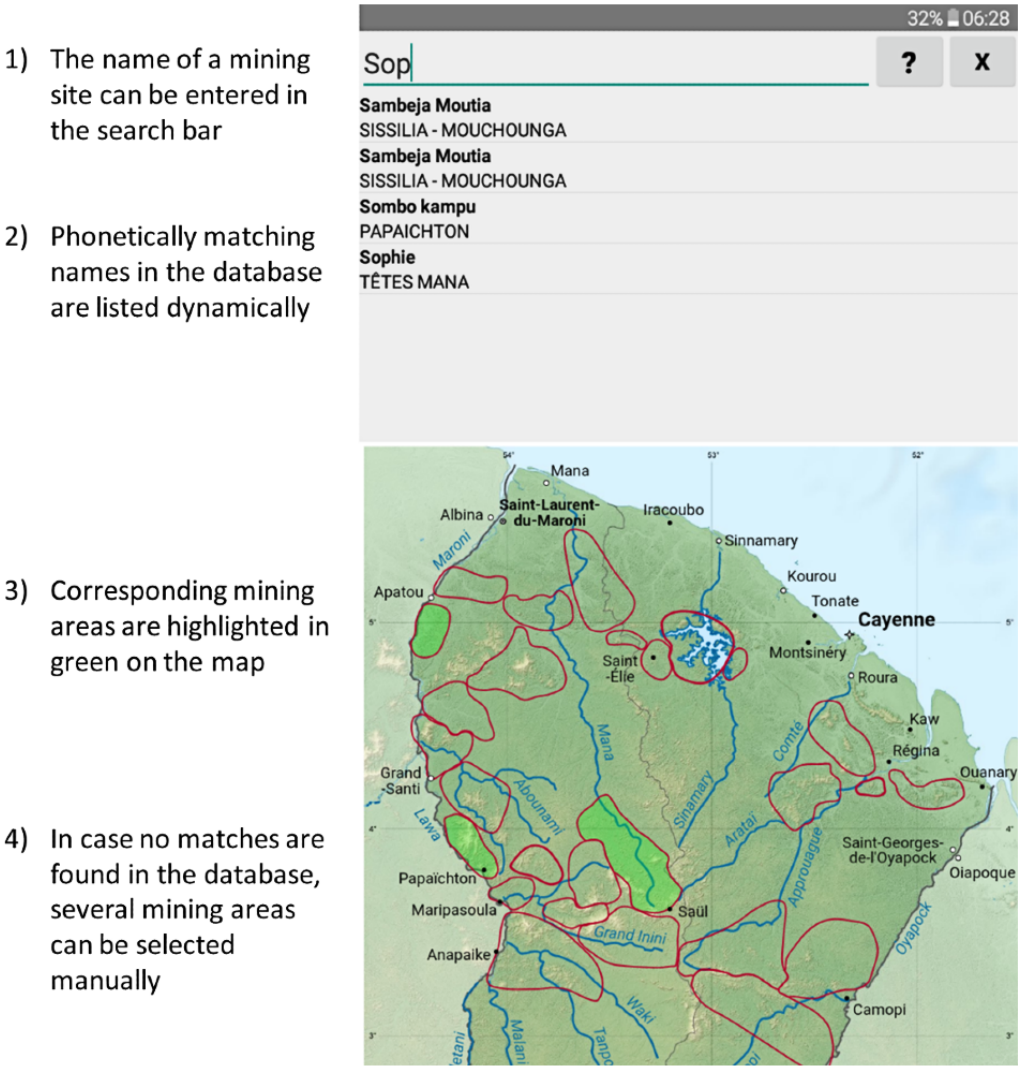
Drawing flowcharts for each questionnaire allowed for a first round of proofreading by explicitly representing skip logic and branching between questions (Multimedia Appendix 1). These also served as valuable references for testing and debugging the forms' behavior in ODK Collect. Online form builders provided by ODK or Ona were not sophisticated enough to address the complexity of the Malakit questionnaires. In this case, documentation from ODK advises designing the form manually in a spreadsheet using the XLSForm (XLS: Microsoft Excel spreadsheet) standard [18,19]. We went a step further and developed a homemade XLSForm template that decouples the form's content from its structure and logic. This facilitated modifications and the eventual translation of all text fields (ie, hints, warnings, questions, and choices for answers) into Portuguese, the main language spoken by the participants (Multimedia Appendix 2).

Malakit facilitators captured participants' information using ODK Collect with the appropriate inclusion or follow-up questionnaire. On inclusion, the participants received a Malakit card with a barcode matching the anonymous ID number on their informed consent form (ICF); the barcode was then scanned with ODK Collect during the interview. In case participants returned for a follow-up visit without a card, facilitators filled in a confidential document with the name and date of birth of each participant and delivered a new anonymous ID number. Versions of ODK Collect ranging from version 1.14 (April 2018) to version 1.18 (November 2018) were used throughout the study. Critical features of ODK Collect, such as form record deletion, were restricted with an administrator password known only by the supervisors. Editing a record after closing and validating a form was impossible because of encryption.

Onde: A Custom App to Capture Names and Locations of Mining Sites

Documenting the names and locations of mining sites was important for improving knowledge about the study population's mobility and for identifying malaria hot spots. However, this was challenging since both participants and facilitators were not familiar with the geography of French Guiana and could use different names for the same site, with different possible spellings [5]. A simple, ad hoc Android app named Onde was developed with a database of names and coordinates of known gold mining sites curated by the Parc Amazonien de Guyane and a map of French Guiana displaying the main rivers, villages, cities, and main gold mining areas. This app aimed to improve user experience and data capture by implementing a phonetic-matching algorithm into an autocompleting search bar synced with an interactive map (Figure 4).

Figure 4. Screenshot of the Onde app (Malakit study, Guiana Shield, April 2018-March 2020).



Data Retrieval, Decryption, and Aggregation: MalakitR Package

In order to streamline data retrieval and aggregation and allow continuous data monitoring, a custom MalakitR package was developed in-house to implement all steps in a single environment using R (versions 3.4.4 to 4.0.0; The R Foundation) and RStudio (versions 1.1.442 to 1.3.959) [20-22].

Encrypted data needed to be downloaded from Ona and then decrypted with ODK Briefcase [23,24]. However, the initial complexity of the follow-up visit questionnaire was improperly handled. In addition, the only solution to re-encrypt data that had been reshaped into an updated form structure and upload them to Ona was to manually re-enter the data into ODK Collect, since ODK Briefcase only handled decryption and not encryption. To implement form encryption and decryption, open-source Java code was extracted from ODK Briefcase (ie, decryption classes) and ODK Collect (ie, encryption classes) and compiled into a single executable decipher.jar file included in the MalakitR package [25,26].

The MalakitR package made it possible to download individual encrypted record files with the help of Ona's application programming interface. Encrypted files that were newly downloaded were decrypted and linearized as single-record files according to the form structure; they were then finally aggregated with existing data to generate an updated table of raw data. During the study, all encrypted and decrypted data were securely stored on the study sponsor's server in French Guiana. The source code of the MalakitR package is available on GitHub [20,25].

Data Monitoring, Validation, and Cleaning

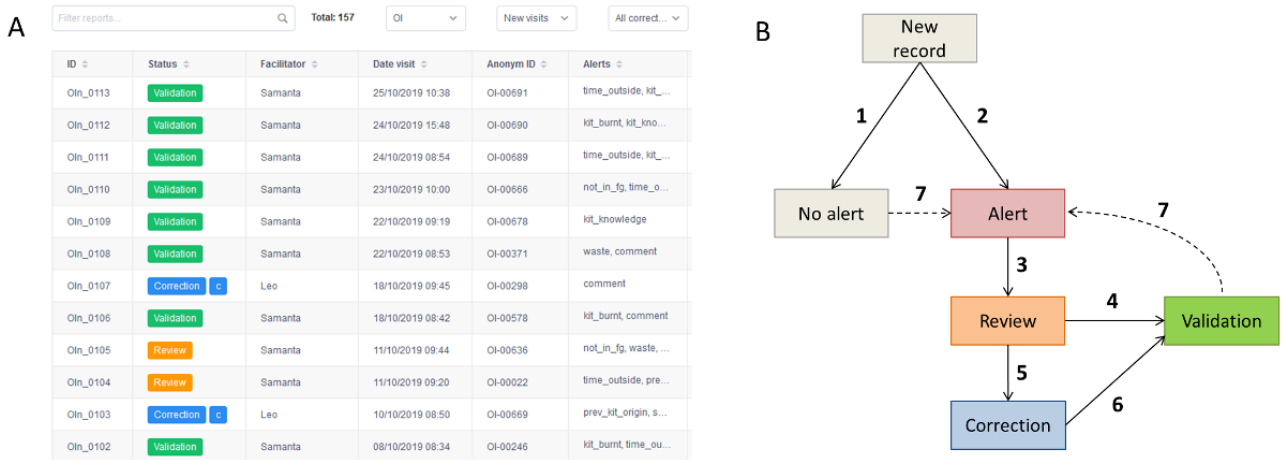
Malakit data collected on the field needed to be monitored continuously in order to ensure their quality and to guarantee that the study protocol was respected. Any issues regarding the safety of participants were to be reported to principal investigators and to the study's Data and Safety Monitoring Board. Because any difficulty or uncertainty was best debriefed with facilitators and local supervisors directly on the field, a simple portable and offline solution was needed.

A portable dashboard was developed as a JavaScript single-page app using the Vue.js framework (version 2.5) [27]. The MalakitR package included an additional set of functions to screen data records against alert rules defined by the coordination team, to detect data worth monitoring manually,

and to generate visit reports that could be read inside the Malakit dashboard.

In the dashboard web app, visit reports displayed pairs of variable names and values in a user-friendly structure. Variables flagged with alerts were highlighted in red to facilitate the review of the data and consequent decision-making. Figure 5 shows a screenshot of the dashboard main screen and the progression of new visit records through the different statuses of the data monitoring process.

Figure 5. Data monitoring with the Malakit dashboard web app (Malakit study, Guiana Shield, April 2018-March 2020) A. Example of data records monitored. B. Flow of data reviewing and validation. Screened visit records are flagged with "No alert" (1) or "Alert" status (2), leading to manual review (3). After review and the facilitator's feedback, records are validated (4) or flagged for correction (5) and patched (6). Records can be screened against new alert rules (7).



Training and Supervision of Facilitators

Practical training sessions were organized throughout the deployment phase of the project. Nine facilitators were trained to fill in the ICF, to assign a new anonymous ID number to a participant, to use Android tablets, to enter data during mock visit interviews with ODK Collect and the Onde app, to upload data using a Wi-Fi network, and to ensure the successful transfer of all pending form records. Training tablets configured with a "sandbox" Ona account were used for this purpose.

Facilitators were supervised on the field at the opening of a new distribution site. Tablets were provided with foldable protective covers and waterproof bags for storage and transportation. Inclusion and follow-up visit forms were specific to each distribution site. This allowed for getting quick activity feedback on the number of forms uploaded from each distribution site directly from Ona. This also avoided multiple updates of the forms in case of facilitator turnover at a single distribution site. Subsequent supervision visits were carried out regularly at all distribution sites by members of the Malakit coordination team and field supervisors. Feedback from facilitators was collected in a nonstructured way throughout the Malakit study during field visits and WhatsApp conversations. At the end of the study, seven facilitators answered a series of questions about their whole experience in the study in the form of short videos, which were displayed during the final meeting of the Malakit project in October 2020 [28].

Ethics Approval

The Malakit study obtained all competent ethical approval required according to the respective regulations in Brazil and Suriname: the ethical committee of the Fundação Oswaldo Cruz in Brazil (CAAE [Certificado de Apresentação de Apreciação Ética] No. 89482118.0.0000.5248; approval No. 2.831.534) and the ethical committee in Suriname (CMWO [Commissie voor Mensgebonden Wetenschappelijk Onderzoek] approval No. VG 25-17). Written consent was obtained from all participants. This study was registered at ClinicalTrials.gov (NCT03695770).

Results

Overview

From April 2018 to March 2020, nine facilitators generated a total of 4863 form records at the five distribution sites of the Malakit project. Herein, we present feedback results from study facilitators and data monitoring by the coordination team.

Feedback From Facilitators

Facilitators' feedback during training and supervision visits at distribution sites was valuable for adapting the information system once the study was fully implemented.

Adaptations Made After Facilitators' Training and Study Launch

Using the tablet, switching Wi-Fi on and off, playing training videos, and using ODK Collect was straightforward enough, since most facilitators were already familiar with Android devices. Removing the answer from an already-completed field in ODK Collect was not intuitive because it required a long press on the field to erase. Learning to use the Onde app was a difficult step of the training for some facilitators who were unfamiliar with the geography of French Guiana or the names of mining sites; however, the facilitators found the interface of the app intuitive enough.

Some adaptations were made to improve the facilitators' work: installation of a specific keyboard app for one facilitator with

sight impairment, use of capital letters instead of digits for months when writing down dates of visit and dates of birth on paper ICFs, and additional practical training for older facilitators who were less experienced in using tablets.

Facilitators were able to clearly explain the context and objectives of the Malakit study to the gold miners, and obtaining their written consent was not an issue. At first, filling in the ICFs correctly with participants was more challenging for facilitators without administrative experience in research or health work. With a little practice, this task quickly became routine, and ICF audits were satisfactory. Facilitators initially confused follow-up visits with inclusion visits on several occasions; this required important clarifications about the difference between the data collected for each type of visit and their purpose in the evaluation of Malakit.

Technical Difficulties

The coordination team in Cayenne provided facilitators with assistance through instant messaging on WhatsApp. Facilitators sent pictures or videos to demonstrate the problem, thereby improving its analysis and resolution. Several technical problems were reported by the facilitators (Table 2), but no big failure of EDC was experienced on the field during the 2 years of the study. In particular, facilitators took great care of the tablets and none were lost, stolen, or damaged, nor showed eventual battery issues.

Table 2. Technical issues with electronic data capture reported by facilitators during the Malakit study, Guiana Shield, April 2018-March 2020.

Category and problem	Frequency	Note	Solution
Hardware			
Loss or damage of the cable and charger of the tablets	Frequent	This happened at all distribution sites.	These were easy to replace with a compatible USB charger and cable.
Slight loss of contrast on the edges of the screen of the tablets	Twice (end of the study)	This was observed at the distribution sites with the highest humidity.	This did not alter the proper functioning of the tablets.
Fear of losing the tablet, leading to use of a blank piece of paper and postponed data entry	Once	This was detected in the dashboard because the kit ID was consistently entered manually and visits were concentrated within a short period of time at the end of a workday.	The facilitator was reassured and asked to use the tablet to avoid errors.
Data entry			
Forgetting to close and validate the form in ODK[a] Collect right at the end of the visit	Frequent (start of the study)	This resulted in a wrong end time stamp and an abnormally long duration of the questionnaire.	Facilitators were shown how to ensure that a form was closed and encrypted. When necessary, data were retrieved manually.
Difficulty scanning kit barcodes in dim light, leading to manual data entry	Frequent	The transparent labels with the kit barcode were pasted on the pink cover of the medication pouch to facilitate the identification of kits, but this resulted in a lower contrast when scanning.	Facilitators were offered to paste the barcode label on a white area inside the kit, at the expense of having to open the kit to scan and identify its ID during inventory.
Scanning barcode of a kit ID instead of an anonymous ID, which blocked the form progression	Once	The form progression was blocked because an incorrect ID format was detected by the regular expression validation constraint. The facilitator simply started again with a new form.	Facilitators were reminded of how to remove an answer in ODK Collect with a long press.
Tablet configuration			
After a prolonged period without use, one tablet was completely discharged and the date was set to the year 1923, blocking the form progression	Once	The form progression was blocked because the automated calculation of the participant's age based on date of birth and the year 1923 returned a negative value.	Facilitators were asked to check that the date of the tablet was correct before launching ODK Collect, especially after a long period without use.
Incorrect time in the tablet due to wrong time zone setting	Once	The time in Suriname, French Guiana, and the North Region of Brazil is GMT–3 throughout the year; in the tablets, this coincided with Buenos Aires instead of Brasilia time, which had daylight time change in February and November (until 2019).	Facilitators were asked to change the time zone setting of the tablet.

[a]ODK: Open Data Kit.

Feedback From Data Monitoring

Data were downloaded and monitored on a weekly basis. WhatsApp messaging was, again, the fastest way of getting feedback from facilitators, but in-depth debriefing was only possible in a face-to-face context on the field. Sometimes a lag time elapsed between data monitoring and actual field supervision, resulting in memory issues when discussing with facilitators. Once facilitators validated a form record, encryption prevented them from looking at the record and editing the data.

In case of doubt regarding the data entered for a particular visit, facilitators were asked to warn the coordination team about the suspected issue for verification. The dashboard was a good tool for visualizing all the information of a visit at once and for guiding the debriefing with facilitators. Supervision visits were more frequent during the months following the opening of a new distribution site or when training of a new facilitator was necessary.

From April 2018 to March 2020, facilitators from the Malakit project generated a total of 4863 form records, corresponding to 3897 inclusion visits (n=2454, 62.97% from Suriname and n=1443, 37.03% from Brazil) and 966 follow-up visits (n=709, 73.4% from Suriname and n=257, 26.6% from Brazil). This corresponded to an average of 202 record submissions per month to the server. Failure to send data was very rare (42/4863, 0.86%): 35 records (0.72%) were retrieved manually from the facilitators' tablets during regular supervision visits at the distribution sites and 7 (0.14%) could not be retrieved. Duplicate records were found for 7 (0.14%) visits. A final number of 4849 unique visits were exploitable with complete data (n=3888, 80.18% inclusion visits and n=961, 19.82% follow-up visits).

Follow-up visits were rare during the first 6 months of the study, but their frequency increased steadily over the following 6-month periods (Figure 6). This was expected, given the median length of stay on mining sites of about 3 months [3]. Sufficient feedback on the follow-up visits was obtained between July and November 2018 to develop a better follow-up questionnaire and a methodology to structure the visit process. The new version of the questionnaire was implemented at all distribution sites in January 2019 during two group training sessions of facilitators. Following this major change in the structure of data, 161 existing follow-up visit records were converted, encrypted, and reuploaded to Ona to ensure the centralization of data and the consistency of the total number of form records submitted.

Data capture with ODK Collect was swift according to the record metadata: the median duration was 5 (IQR 3-7) minutes (n=4856), with both types of visits combined. Given that visits usually lasted an estimated total of 30 to 45 minutes, this suggests that EDC was not taking more time from the training or debriefing of participants. A minority of 1.32% (64/4856) of the durations exceeded 60 minutes, usually because the form records were not closed and validated immediately at the end of an interview. For inclusion questionnaires, a median duration of 4 (IQR 3-6) minutes (n=3892) was measured, with a median number of 27 (IQR 26-28) variables completed. The median duration was similar for the follow-up questionnaire when the participants did not report malaria episodes: 5 (IQR 3-7) minutes (n=611) for a median number of 24 (IQR 23-25) variables completed. The follow-up questionnaire became lengthier when participants reported at least one episode of malaria: median duration was 10 (IQR 7-15) minutes (n=350) for a median number of 46 (IQR 41-50) variables completed. Figure 7 shows that the duration of the questionnaires decreased over the course of the study. This was especially true for inclusion visits: 62.02% (1246/2009) versus 34.09% (642/1883) of inclusion questionnaires lasted more than 4 minutes during the first and second versus third and fourth 6-month periods of the study, respectively (χ^2_1=303.5, P<.001). This suggests that facilitators experienced a learning curve while using ODK Collect.

Figure 6. Number and proportion of records for inclusion and follow-up visits according to the 6-month period of study (Malakit study, Guiana Shield, April 2018-March 2020).

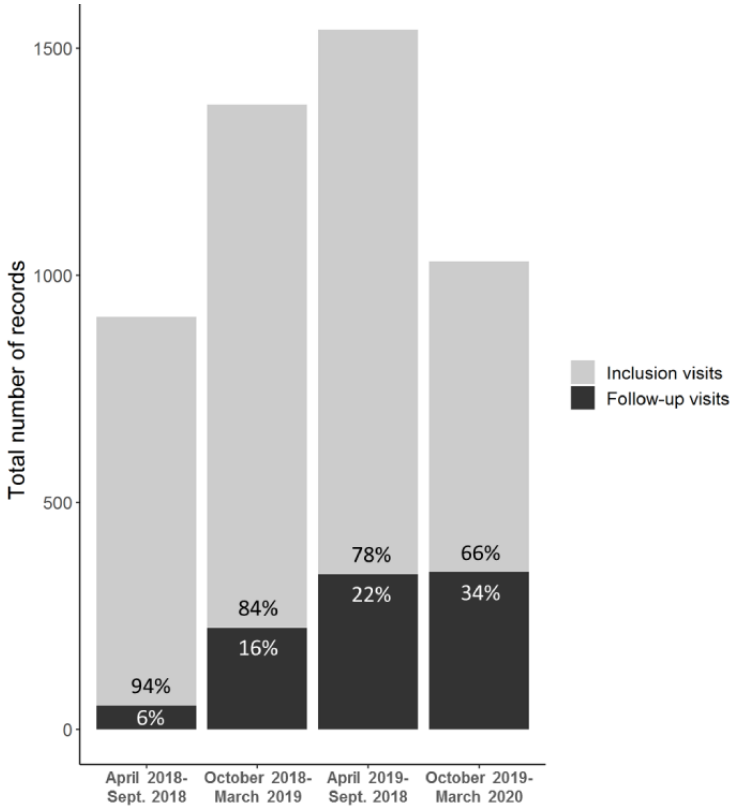
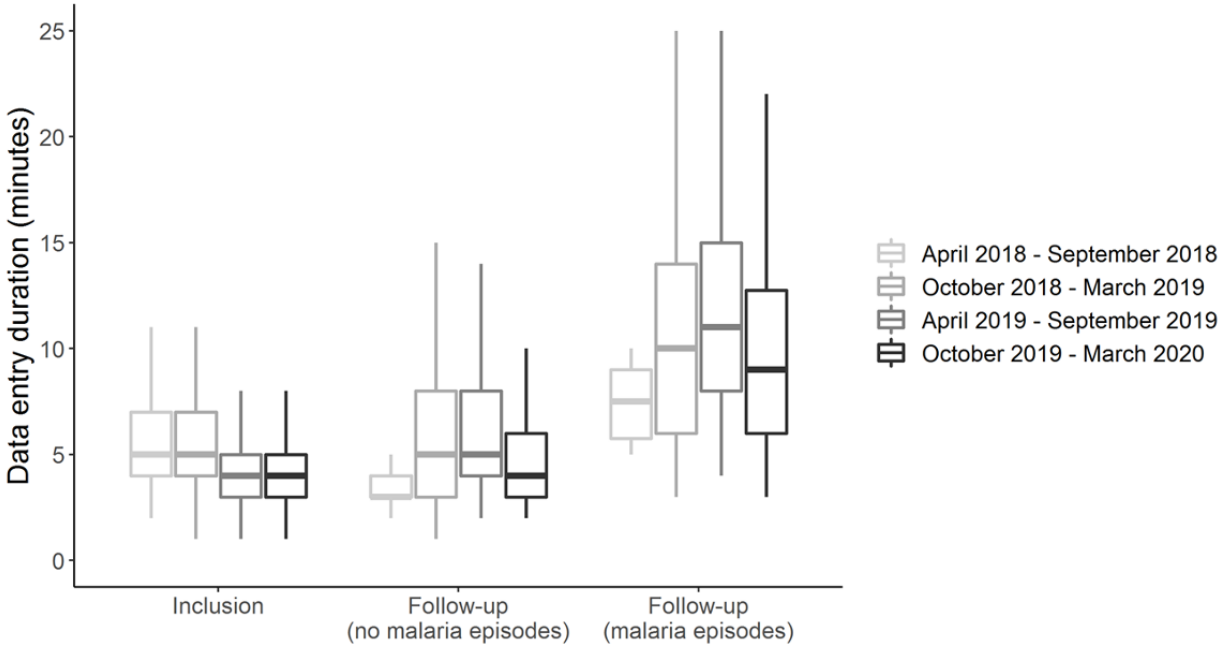


Figure 7. Duration of data entry for the inclusion and follow-up questionnaires, with and without malaria episodes reported by participants (Malakit study, Guiana Shield, April 2018-March 2020). The horizontal lines within the boxes represent medians and the whiskers represent IQRs.



Over the course of the study, 18.31% (712/3888) of inclusion visits and 90.6% (871/961) of follow-up visits raised alerts; the discrepancy results from the greater number of alerts configured specifically for follow-up visits due to their higher complexity. A large proportion of follow-up visits (381/961, 39.6%) included written comments from facilitators. This confirmed the relevance of this feature to collect facilitators' feedback and improve the quality of data collection and supervision. In total, 14.23% (690/4849) of visits raised alerts related to data entry or data accuracy. The most common alerts for data entry were manual barcode entry (158/4849, 3.26%), visit duration longer than 60 minutes (112/4849, 2.31%), and incoherent date of visit (34/4849, 0.70%). An important proportion of visits required feedback from facilitators either from WhatsApp or field supervision visits: 20.4% (145/712) of inclusion visits and 32.0% (279/871) of follow-up visits with alerts. Corrections were made to the data for 3.03% (147/4849) of the total visits, more frequently for follow-up visits (108/961, 11.2%) than inclusion visits (39/3888, 1.00%). Problems with data quality or protocol compliance were more frequent at the two most remote distribution sites. This could be linked to the higher difficulty for the coordination team to travel to these sites on a regular basis.

In 37.7% (362/961) of follow-up visits, participants were able to show an anonymous ID card. At the end of the ICF audits, 52 out of 3888 (1.34%) inclusion visits were found without a documented ICF, mainly owing to the loss of the last batch of 50 paper-based ICFs from Brazil at the end of the study. Although unfortunate, the removal of associated data was of limited impact because it mostly concerned participants included at the end of the study and very unlikely to attend a follow-up visit.

Comparison of dates of birth of the participants both written in the ICFs and entered in ODK Collect by facilitators provided insight into the consistency of data entry. Among the 3777 ICFs with a known date of birth (52 missing and 59 with an estimated age only), 113 (2.99%) mismatches with form records were identified, 73.5% (83/113) of which were minor differences, such as an increment or a decrement of 1 year.

Discussion

Adapt and Reuse; Do Not Reinvent the Wheel

The information system of the Malakit study is a patchwork of existing open-source services (ie, ODK Collect and R packages) and commercial services (ie, Ona) as well as tools developed in-house, of which some are innovations and may be adapted to other contexts beyond interventional health research. In particular, we developed an Android app compatible with ODK Collect as a means to provide a dynamic user interface and improve the collection of fuzzy geographical data. We also developed an R package allowing data managers to retrieve and clean data captured with ODK Collect and that supports bidirectional encryption and decryption.

The development of the Malakit information system was guided by the needs and constraints of the project. According to the MDC manual of the US Global Development Lab, the design and setting of the Malakit study fit into the highly complex category for all three dimensions examined: survey, analysis, and local context [29]. In this situation, the manual advises the hiring of an external consultant or the development of internal capacity. In order to better adapt to the complex and iterative implementation of the project, we decided that the development of the information system and the data management would rely on in-house human resources and technical skills. We chose, however, to contract with an external company to ensure the quality and compatibility of the Malakit smartphone app, which was designed as an additional evaluation tool of the Malakit project [6]. This app can be considered as a separate information system and mHealth intervention; its complete description and evaluation will be the object of a separate paper.

Our strategy was to reuse existing mature technology and fill the gaps with custom and retro-engineered components adapted to specific needs. In the event that in-house development of new tools is an option, as in this project, one should be careful not to reinvent the wheel and should consider existing tools and focus on developing tools specifically tailored to the needs of the project [30,31]. In that respect, mature open-source ecosystems, such as ODK, allow for a trade-off between innovation, adaptation, and reuse of existing solutions. ODK Collect, along with ODK Aggregate and ODK Briefcase, has been widely used in other settings and modified to answer particular needs in some studies [32-38]. The use of R or Stata downstream of ODK has been implemented by other research teams with success; the MalakitR package is another contribution to this open ecosystem [32,39]. The portable dashboard was a prototype that will not be maintained because it relies, in part, on deprecated technology. However, it helped specify requirements for the external development of a similar tool for future field studies.

Safety and confidentiality of personal data are essential and should dissuade nonprofessional developers to build their own system from scratch, especially when online transfer of health-related data is at stake [40,41]. Open or interoperable alternatives to ODK, such as REDCap (Research Electronic Data Capture), Enketo, CommCare, or OpenMRS (medical record system), may be worth considering for data collection and monitoring [31,42-45]. No unique EDC solution would fit all the needs of a study; some authors even recommend using several EDC tools when relevant [46]. Selecting components of an EDC system is not easy, and both field experience and structured guidance are valuable [29,47,48]. Provided that frequent updates are made to keep up with the rapid evolution of technology, online decision tools, such as NOMAD (Humanitarian Operations Mobile Acquisition of Data) or Kopernik, may help in specifying one's needs and in finding an appropriate solution [49,50].

Strengths and Weaknesses: Lessons Learned From the Field

Our experience is another contribution to the evidence reported in the literature, in that health mediators can successfully collect complex data using MDC tools when provided with adequate training and supervision [51]. Training facilitators, first in groups and then individually on the field, proved beneficial to better adapt the MDC strategy and increase their confidence with the technology they used. This experience is in line with the conclusions and recommendations from other studies [31,36,38,52,53]. Field staff turnover occurred at several distribution sites during the 2 years of the study and required timely training sessions, including, but not limited to, data collection. On one occasion, the training was successfully delivered by an already-experienced facilitator, allowing the coordination team to later focus on building capacity on specific points in the questionnaires.

The information system was reliable and the quality of the data collected by facilitators was satisfactory, allowing for a robust analysis of the main results of the study [54]. The choice of tablets was eventually strengthened by the absence of loss, damage, or battery failure, unlike that experienced in other studies [39]. A 7-inch screen, larger than a smartphone, allowed for the display of training material and was of great help for one sight-impaired facilitator, as reported by Dickinson et al [55]. Few technical issues were reported, few visits failed to be sent to the server, and virtually no visits lacked exploitable data. A significant number of visits required feedback from facilitators, but very few resulted in data correction. The audits of the ICFs found limited and minor mistakes in the data entry of dates of birth. Remote and field debriefings with facilitators aided by a portable dashboard web app greatly improved the monitoring of data and the detection of protocol issues or difficulties with the questionnaires. Feedback from the first follow-up visits helped bring major modifications to the follow-up questionnaire with a quick implementation, resulting in the integration of free comments from facilitators, which, in turn, helped with debriefing and contextualizing their issues.

Form testing is an essential precaution noted by other authors [39,46,52,56]. Knowing that losing any of the gold miners' time would be detrimental to their participation in the study, we put a lot of effort into testing and adapting the electronic forms according to the facilitators' feedback. This optimized the forms' design and usability. Representing the branching logic of questions with a flowchart was useful for clarifying the expected electronic behavior of questionnaires that were drafted on paper and for guiding test simulations. Building an XLSForm template also sped up translations and modifications of the form. All of these efforts contributed to the short duration of data entry by facilitators, as shown by the analysis of the form metadata generated by ODK Collect. This freed up time for the actual training or debriefing of participants during visits.

The advantages of EDC over paper questionnaires are well established, especially in the case of complex studies [34,55,57]. EDC suppresses the need for long and error-prone data entry and, thus, shortens the delay between data collection and analysis [58]. However, EDC may in some situations raise issues of trust between data collectors and participants [35,58]. In Malakit, the careful hiring of facilitators from the population of interest helped in that regard; trust of participants in both the facilitators and the investigators was key to the success of the intervention [7]. Still, paper was used to collect informed consent and was, thus, not totally eliminated from the Malakit study. The loss of one batch of 50 ICFs illustrates the hazards associated with a fully paper-based system, especially in a complex setting such as ours. Although the removal of associated data had no impact on the outcome measure (ie, correct kit use) and virtually all ICFs were retrieved and audited successfully, electronic alternatives may be worth implementing for future projects [59,60].

The use of anonymous ID cards was not efficient. Few participants came back with their card, and mix-ups between participants who had attended the same inclusion visit were occasionally detected in the data. Participants of the study trusted the facilitators with their identity and did not have a problem disclosing it again when they could not show a card. Such challenges in identifying study participants from hard-to-reach populations have been reported with ID cards [61]. Alternatives such as fingerprints or photographs may be

considered, but their acceptability in the gold miner population would require careful participatory evaluation and validation, as well as ethical and regulatory clearance [32,61].

Perspectives

The different tasks associated with the Malakit information system fit into a continuum between digital tools development, field supervision, and data management, which all require different levels of technical skills and time resources. These tasks could be achieved by a single person during the 2-year period of the study, in part owing to the low volume of data generated in our study when compared with other studies: less than 4000 participants were included in Malakit versus 25,000 and 280,000 participants according to Style et al and Marks et al, respectively [32,39]. For such larger studies, a team dedicated to data management, supervision, and system development should be set up [41,45]. One lesson learned from the field was the need for local supervisors fully dedicated to the research project. A solution to maintain more constant and reactive data supervision would be to delegate part of the process to local supervisors through capacity building [58].

Strategies to improve data collection directly from the mining areas could be evaluated using text messaging, instant messaging, or ODK Collect with Bluetooth data transfer [38,62]. The dashboard web app used for the Malakit study was a prototype that would only suit the specific needs of the supervision team. A new dashboard developed as a stand-alone Shiny app would better interface with the MalakitR package [63]. Setting up our own instance of ODK Central would help in complying more easily with French and European regulations regarding data protection and in using the R package ruODK as an interface [32,64]. In the long run, ODK-X offers interesting perspectives for the management of longitudinal data [65,66].

The components of the Malakit information system can be useful as a whole or in part for research teams wishing to collect and monitor encrypted data in similar conditions as in our study: (1) remote inclusion sites with little access to internet and electricity, (2) data collected by people without experience in research, and (3) continuous data monitoring and quality control. The system can be implemented and managed by personnel with sufficient computer skills and knowledge of R, in close coordination with field workers and supervisors.

The information system described in this paper evolved into a simplified version for monitoring Malakit as a public health intervention in Suriname. It may contribute to a coherent and coordinated public health action in the region and eventually interface with existing cross-border surveillance systems of malaria [67].

Conclusions

The development of the information system for the Malakit project was a source of innovation that mirrored the inventiveness of the intervention itself. Unprecedented in this particular context, the development had to adapt iteratively to the constraints and different phases of the study. Some of the components are available for reuse in different research contexts. Our experience with the Malakit information system confirms that, provided that adequate care is brought to design, training, and supervision, health mediators are able to capture good-quality data to support the evaluation of a complex research study in a challenging environment.

Acknowledgments

The authors thank all facilitators from the Malakit project. Facilitators in Suriname include the following: Breno Otavio Barbosa, Isadora Batista dos Santos, Astrid Huur, Carla Pereira de Paulo, Fernando Silva Rodrigues Barbosa, and Nicholas Valente Ferreira. Facilitators in Brazil include the following: Samanta Braga, Salomao Brito, Alejandro Chavarro, André Da Silva Alcântara, Ronnie Elison Alves Sampaio, and Leonardo Villalba-Bassulto. The authors also thank Pierre Joubert (Parc Amazonien de Guyane); Hélène Hilderal, Aniza Fahrasmane, and Amandine Duclau (Cayenne Hospital); and the Open Data Kit community. Pictograms used in figures were designed by Lorena Salagre, Yo! Baba, Brennan Novax, Adnen Kadri, il Capitano, and Gregor Cresnar from the Noun Project. The R logo was reused according to the CC BY-SA 4.0 (Creative Commons Attribution-ShareAlike 4.0 International) license. The map of French Guiana included in the Onde app was adapted from Sémhur, Wikimedia Commons, according to the CC BY-SA 3.0 (Creative Commons Attribution-ShareAlike 3.0 Unported) license.

Cayenne Hospital, as a study sponsor, directly managed most of the project funds. Half of the funding was provided by the European Regional Development Fund via the Interregional Amazon Cooperation Program 2014-2020. This was supplemented with self-funding from Cayenne Hospital, funds from the French Guiana Health Regional Agency, funds from the European Horizon 2020 Program, and in-kind contributions from the Suriname National Malaria Control Programme, as beneficiary of the Global Fund to Fight AIDS, Tuberculosis and Malaria; the Ministry of Health of Brazil; and the French Development Agency (Agence Française de Développement). The funding bodies played no part in the implementation and evaluation of the project.

Authors' Contributions

MD, MG, YL, MSM, SV, MN, AA, HH, PM, and AS conceived the study and wrote the protocol. YL, MG, LG, and MD designed and developed the information system. MG, YL, and LM performed field implementation and supervision and collected feedback with the help of JBM, HC, JHG, and SV. YL and MD wrote the first draft of the manuscript. MG, MSM, SV, AS, HH, AA, and MN reviewed the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Flowchart showing skip logic and branching of inclusion questionnaire.
[PDF File (Adobe PDF File), 130 KB - formative_v6i6e29856_app1.pdf]

Multimedia Appendix 2
Template XLS form used for inclusion questionnaire. XLS: Microsoft Excel spreadsheet.
[XLS File (Microsoft Excel File), 589 KB - formative_v6i6e29856_app2.xls]

References

1. Douine M, Lambert Y, Musset L, Hiwat H, Blume L, Marchesini P, et al. Malaria in gold miners in the Guianas and the Amazon: Current knowledge and challenges. Curr Trop Med Rep 2020 Mar 21;7(2):37-47. [doi: 10.1007/s40475-020-00202-5]
2. Douine M, Musset L, Corlin F, Pelleau S, Pasquier J, Mutricy L, et al. Prevalence of Plasmodium spp. in illegal gold miners in French Guiana in 2015: A hidden but critical malaria reservoir. Malar J 2016 Jun 09;15(1):315 [FREE Full text] [doi: 10.1186/s12936-016-1367-6] [Medline: 27277831]
3. Douine M, Mosnier E, Le Hingrat Q, Charpentier C, Corlin F, Hureau L, et al. Illegal gold miners in French Guiana: A neglected population with poor health. BMC Public Health 2017 Jul 17;18(1):23 [FREE Full text] [doi: 10.1186/s12889-017-4557-4] [Medline: 28716015]
4. Douine M, Lazrek Y, Blanchet D, Pelleau S, Chanlin R, Corlin F, et al. Predictors of antimalarial self-medication in illegal gold miners in French Guiana: A pathway towards artemisinin resistance. J Antimicrob Chemother 2018 Jan 01;73(1):231-239. [doi: 10.1093/jac/dkx343] [Medline: 29045645]
5. Le Tourneau FM. Chercheurs d'Or - L'Orpaillage Clandestin en Guyane Française. Paris, France: CNRS Éditions; Jul 02, 2020.
6. Douine M, Sanna A, Galindo M, Musset L, Pommier de Santi V, Marchesini P, et al. Malakit: An innovative pilot project to self-diagnose and self-treat malaria among illegal gold miners in the Guiana Shield. Malar J 2018 Apr 10;17(1):158 [FREE Full text] [doi: 10.1186/s12936-018-2306-5] [Medline: 29631588]
7. Galindo MS, Lambert Y, Mutricy L, Garancher L, Bordalo Miller J, Gomes JH, et al. Setting-up a cross-border action-research project to control malaria in remote areas of the Amazon: Describing the birth and milestones of a complex international project (Malakit). Malar J 2021 May 11;20(1):216 [FREE Full text] [doi: 10.1186/s12936-021-03748-5] [Medline: 33975624]
8. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Council guidance. BMJ 2015 Mar 19;350(mar19 6):h1258 [FREE Full text] [doi: 10.1136/bmj.h1258] [Medline: 25791983]
9. Galindo MS, Lambert Y, Mutricy L, Garancher L, Miller JB, Gomes JH, et al. Implementation of a novel malaria management strategy based on self-testing and self-treatment in remote areas in the Amazon (Malakit): Confronting a-priori assumptions with reality. BMC Public Health 2022 Apr 15;22(1):770 [FREE Full text] [doi: 10.1186/s12889-022-12801-0] [Medline: 35428230]
10. Mosnier E, Garancher L, Galindo M, Djossou F, Moriceau O, Hureau-Mutricy L, et al. Paludisme en Guyane: Des projets de recherche opérationnelle originaux s 'appuyant sur la santé communautaire. Lett Infect 2020 Apr 30:50-57.
11. The kit. Malakit Project. URL: https://www.malakit-project.org/kit/ [accessed 2022-04-28]
12. Galaxy Tab A 7.0. Samsung. URL: https://www.samsung.com/us/business/support/owners/product/galaxy-tab-a-7-0-wi-fi/ [accessed 2022-04-28]
13. Collect data anywhere. ODK. URL: https://getodk.org [accessed 2022-04-28]
14. Hartung C. Open Data Kit: Technologies for Mobile Data Collection and Deployment Experiences in Developing Regions [doctoral thesis]. Seattle, WA: University of Washington; 2012 Sep 13. URL: https://digital.lib.washington.edu:443/researchworks/handle/1773/20748 [accessed 2020-12-21]
15. General Data Protection Regulation (GDPR). URL: https://gdpr-info.eu/ [accessed 2022-04-28]
16. Ona. URL: https://company.ona.io/ [accessed 2022-04-28]
17. KoBoToolbox. URL: https://kobotoolbox.org/ [accessed 2022-04-28]
18. Form building tools. ODK Docs. URL: https://docs.getodk.org/form-tools/ [accessed 2022-04-28]
19. XLSForm.org. URL: https://xlsform.org/en/ [accessed 2022-04-28]
20. Lambert Y. MalakitR package. GitHub. URL: https://github.com/ymlambert/malakitr [accessed 2022-04-28]
21. The R Project for Statistical Computing. R Project. URL: https://www.r-project.org/ [accessed 2022-04-28]
22. RStudio. URL: https://rstudio.com/ [accessed 2022-04-28]
23. Ona JSON REST API endpoints. Ona API. URL: https://api.ona.io/static/docs/index.html [accessed 2022-04-28]
24. ODK Briefcase. ODK Docs. URL: https://docs.getodk.org/briefcase-intro/ [accessed 2022-04-28]

25. Lambert Y. Decipher. GitHub. URL: https://github.com/ymlambert/decipher [accessed 2022-04-28]
26. Security and privacy. ODK Docs. URL: https://docs.getodk.org/security-privacy/ [accessed 2022-04-28]
27. Vue.js. URL: https://vuejs.org/ [accessed 2022-04-28]
28. The facilitators: Key people in the Malakit project. Malakit Project. 2020 Oct. URL: https://www.malakit-project.org/the-facilitators-key-people-in-the-malakit-project/ [accessed 2022-04-28]
29. Satterlee E, McCullough L, Dawson M, Cheung K. Paper-to-Mobile Data Collection: A Manual. Washington, DC: US Global Development Lab, USAID; 2018 Oct 18. URL: https://www.fhi360.org/sites/default/files/media/documents/Paper_to_Mobile_Data_Collection_Manual_1.0.pdf [accessed 2022-04-30]
30. Park JY, Kim DR, Haldar B, Mallick AH, Kim SA, Dey A, et al. Use of the data system for field management of a clinical study conducted in Kolkata, India. BMC Res Notes 2016 Jan 09;9:20 [FREE Full text] [doi: 10.1186/s13104-015-1767-7] [Medline: 26749186]
31. Labrique AB, Wadhwani C, Williams KA, Lamptey P, Hesp C, Luk R, et al. Best practices in scaling digital health in low and middle income countries. Global Health 2018 Nov 03;14(1):103 [FREE Full text] [doi: 10.1186/s12992-018-0424-z] [Medline: 30390686]
32. Marks M, Lal S, Brindle H, Gsell PS, MacGregor M, Stott C, et al. Electronic data management in public health and humanitarian crises. Upgrades, scalability and impact of ODK. Research Square. Preprint posted online on August 12, 2020 [FREE Full text] [doi: 10.21203/rs.3.rs-52854/v1]
33. Tom-Aba D, Olaleye A, Olayinka AT, Nguku P, Waziri N, Adewuyi P, et al. Innovative technological approach to Ebola virus disease outbreak response in Nigeria using the Open Data Kit and Form Hub technology. PLoS One 2015 Jun 26;10(6):e0131000 [FREE Full text] [doi: 10.1371/journal.pone.0131000] [Medline: 26115402]
34. Maduka O, Akpan G, Maleghemi S. Using Android and Open Data Kit technology in data management for research in resource-limited settings in the Niger Delta region of Nigeria: Cross-sectional household survey. JMIR Mhealth Uhealth 2017 Nov 30;5(11):e171 [FREE Full text] [doi: 10.2196/mhealth.7827] [Medline: 29191798]
35. Naker K, Gaskell KM, Dorjravdan M, Dambaa N, Roberts CH, Moore DAJ. An e-registry for household contacts exposed to multidrug resistant TB in Mongolia. BMC Med Inform Decis Mak 2020 Aug 12;20(1):188 [FREE Full text] [doi: 10.1186/s12911-020-01204-z] [Medline: 32787837]
36. Sivaraman S, Soni P. Leveraging technology for optimization of health survey research. J Health Manag 2019 Nov 25;21(4):571-581. [doi: 10.1177/0972063419884444]
37. Sharif B, Lundin RM, Morgan P, Hall JE, Dhadda A, Mann C, Welsh Digital Data Collection Platform Collaborators (see Appendix A). Developing a digital data collection platform to measure the prevalence of sepsis in Wales. J Am Med Inform Assoc 2016 Nov;23(6):1185-1189. [doi: 10.1093/jamia/ocv208] [Medline: 27094989]
38. Kenny A, Gordon N, Downey J, Eddins O, Buchholz K, Menyon A, et al. Design and implementation of a mobile health electronic data capture platform that functions in fully-disconnected settings: A pilot study in rural Liberia. BMC Med Inform Decis Mak 2020 Feb 22;20(1):39 [FREE Full text] [doi: 10.1186/s12911-020-1059-6] [Medline: 32087731]
39. Style S, Beard BJ, Harris-Fry H, Sengupta A, Jha S, Shrestha BP, et al. Experiences in running a complex electronic data capture system using mobile phones in a large-scale population trial in southern Nepal. Glob Health Action 2017 Jun 14;10(1):1330858 [FREE Full text] [doi: 10.1080/16549716.2017.1330858] [Medline: 28613121]
40. Chow PI. Developing mental or behavioral health mobile apps for pilot studies by leveraging survey platforms: A do-it-yourself process. JMIR Mhealth Uhealth 2020 Apr 20;8(4):e15561 [FREE Full text] [doi: 10.2196/15561] [Medline: 32310143]
41. Katarahweire M, Bainomugisha E, Mughal KA. Data classification for secure mobile health data collection systems. Dev Eng 2020;5:100054. [doi: 10.1016/j.deveng.2020.100054]
42. Ruth CJ, Huey SL, Krisher JT, Fothergill A, Gannon BM, Jones CE, et al. An electronic data capture framework (ConnEDCt) for global and public health research: Design and implementation. J Med Internet Res 2020 Aug 13;22(8):e18580 [FREE Full text] [doi: 10.2196/18580] [Medline: 32788154]
43. Harris-Fry H, Beard BJ, Harrisson T, Paudel P, Shrestha N, Jha S, et al. Smartphone tool to collect repeated 24 h dietary recall data in Nepal. Public Health Nutr 2017 Aug 31;21(2):260-272. [doi: 10.1017/s136898001700204x]
44. Leal Neto OB, Loyo R, Albuquerque J, Perazzo J, Barbosa V, Barbosa CS. Using mobile technology to conduct epidemiological investigations. Rev Soc Bras Med Trop 2015 Feb;48(1):105-107 [FREE Full text] [doi: 10.1590/0037-8682-0181-2014] [Medline: 25860475]
45. Blumenberg C, Barros AJD. Electronic data collection in epidemiological research. The use of REDCap in the Pelotas birth cohorts. Appl Clin Inform 2016 Jul 13;7(3):672-681 [FREE Full text] [doi: 10.4338/ACI-2016-02-RA-0028] [Medline: 27453034]
46. Comulada WS, Tang W, Swendeman D, Cooper A, Wacksman J, Adolescent Medicine Trials Network (ATN) CARES Team. Development of an electronic data collection system to support a large-scale HIV behavioral intervention trial: Protocol for an electronic data collection system. JMIR Res Protoc 2018 Dec 14;7(12):e10777 [FREE Full text] [doi: 10.2196/10777] [Medline: 30552083]
47. Steinberg M, Schindler S, Klan F. Software solutions for form-based, mobile data collection - A comparative evaluation. In: Proceedings of the Datenbanksysteme für Business, Technologie und Web - Workshop. Gesellschaft für Informatik.

2019 Presented at: Datenbanksysteme für Business, Technologie und Web - Workshop. Gesellschaft für Informatik; March 4-8, 2019; Rostock, Germany p. 135-144 URL: https://dl.gi.de/bitstream/handle/20.500.12116/21824/LNI-P-290-BTW2019-Workshopband.pdf?sequence=1&isAllowed=y [doi: doi:10.18420/btw2019-ws-14]
48. King C, Hall J, Banda M, Beard J, Bird J, Kazembe P, et al. Electronic data capture in a rural African setting: Evaluating experiences with different systems in Malawi. Glob Health Action 2014 Oct 30;7(1):25878 [FREE Full text] [doi: 10.3402/gha.v7.25878] [Medline: 25363364]
49. Humanitarian Operations Mobile Acquisition of Data (NOMAD). Elrha. URL: https://www.elrha.org/project/humanitarian-operations-mobile-acquisition-data-nomad/ [accessed 2022-04-28]
50. Impact Tracker Tech. Kopernik. URL: http://impacttrackertech.kopernik.info/ [accessed 2022-04-28]
51. Braun R, Catalani C, Wimbush J, Israelski D. Community health workers and mobile technology: A systematic review of the literature. PLoS One 2013 Jun;8(6):e65772 [FREE Full text] [doi: 10.1371/journal.pone.0065772] [Medline: 23776544]
52. Shiferaw S, Workneh A, Yirgu R, Dinant G, Spigt M. Designing mHealth for maternity services in primary health facilities in a low-income setting - Lessons from a partially successful implementation. BMC Med Inform Decis Mak 2018 Nov 12;18(1):96 [FREE Full text] [doi: 10.1186/s12911-018-0704-9] [Medline: 30419891]
53. Webster-Smith M, Paulding C, Burnett S, Jeffs L, Ismail D, McNamara C, et al. Implementing electronic data capture (EDC) training for site staff. In: Proceedings of the 3rd International Clinical Trials Methodology Conference. 2015 Presented at: The 3rd International Clinical Trials Methodology Conference; November 16-17, 2015; Glasgow, UK p. P41 URL: https://trialsjournal.biomedcentral.com/track/pdf/10.1186/1745-6215-16-S2-P41.pdf [doi: 10.1186/1745-6215-16-s2-p41]
54. Douine M, Lambert Y, Galindo M, Mutricy L, Sanna A, Peterka C, et al. Self-diagnosis and self-treatment of malaria in hard-to-reach and mobile populations of the Amazon: Results of Malakit, an international multicentric intervention research project. Lancet Reg Health Am 2021 Dec;4:100047. [doi: 10.1016/j.lana.2021.100047]
55. Dickinson FM, McCauley M, Madaj B, van den Broek N. Using electronic tablets for data collection for healthcare service and maternal health assessments in low resource settings: Lessons learnt. BMC Health Serv Res 2019 May 27;19(1):336 [FREE Full text] [doi: 10.1186/s12913-019-4161-7] [Medline: 31133032]
56. Mugisha A, Krumsvik O, Tylleskar T, Babic A. Data collectors' design preferences for mobile electronic data capturing forms. In: Hasman A, Gallos P, Liaskos J, Househ MS, Mantas J, editors. Data, Informatics and Technology: An Inspiration for Improved Healthcare. Amsterdam, the Netherlands: IOS Press BV; 2018:93-96.
57. Ley B, Rijal KR, Marfurt J, Adhikari NR, Banjara MR, Shrestha UT, et al. Analysis of erroneous data entries in paper based and electronic data collection. BMC Res Notes 2019 Aug 22;12(1):537 [FREE Full text] [doi: 10.1186/s13104-019-4574-8] [Medline: 31439025]
58. Stover B, Lubega F, Namubiru A, Bakengesa E, Luboga SA, Makumbi F, et al. Conducting a large public health data collection project in Uganda: Methods, tools, and lessons learned. J Res Pract 2018;14(1):M1 [FREE Full text] [Medline: 30057632]
59. De Sutter E, Zaçe D, Boccia S, Di Pietro ML, Geerts D, Borry P, et al. Implementation of electronic informed consent in biomedical research and stakeholders' perspectives: Systematic review. J Med Internet Res 2020 Oct 08;22(10):e19129 [FREE Full text] [doi: 10.2196/19129] [Medline: 33030440]
60. Skelton E, Drey N, Rutherford M, Ayers S, Malamateniou C. Electronic consenting for conducting research remotely: A review of current practice and key recommendations for using e-consenting. Int J Med Inform 2020 Nov;143:104271 [FREE Full text] [doi: 10.1016/j.ijmedinf.2020.104271] [Medline: 32979650]
61. Kaewkungwal J, Apidechkul T, Jandee K, Khamsiriwatchara A, Lawpoolsri S, Sawang S, et al. Application of mobile technology for improving expanded program on immunization among highland minority and stateless populations in northern Thailand border. JMIR Mhealth Uhealth 2015 Jan 14;3(1):e4 [FREE Full text] [doi: 10.2196/mhealth.3704] [Medline: 25589367]
62. Ullrich L, Khoudary H. WhatsApp Surveying Guide: Lessons Learnt From Two Qualitative WhatsApp Surveys in Lebanon. New York, NY: United Nations Development Programme; 2018. URL: https://data2.unhcr.org/en/documents/details/67580 [accessed 2022-04-30]
63. COVAIL (GoSecure). Photon package. GitHub. URL: https://github.com/ColumbusCollaboratory/photon [accessed 2022-04-28]
64. Mayer FW. ruODK: An R client for the ODK central API. Zenodo. 2021. URL: https://zenodo.org/record/5559164 [accessed 2022-04-28]
65. ODK-X. URL: https://odk-x.org/ [accessed 2022-04-28]
66. Brunette W, Sudar S, Sundt M, Larson C, Beorse J, Anderson R. Open Data Kit 2.0: A services-based application framework for disconnected data management. In: Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services. New York, NY: Association for Computing Machinery; 2017 Presented at: The 15th Annual International Conference on Mobile Systems, Applications, and Services; June 19-23, 2017; Niagara Falls, NY p. 440-452. [doi: http://dx.doi.org/10.1145/3081333.3081365]
67. Saldanha R, Mosnier É, Barcellos C, Carbunar A, Charron C, Desconnets J, et al. Contributing to elimination of cross-border malaria through a standardized solution for case surveillance, data sharing, and data interpretation: Development of a

cross-border monitoring system. JMIR Public Health Surveill 2020 Sep 01;6(3):e15409 [FREE Full text] [doi: 10.2196/15409] [Medline: 32663141]

Abbreviations

CAAE: Certificado de Apresentação de Apreciação Ética
CC BY-SA 3.0: Creative Commons Attribution-ShareAlike 3.0 Unported
CC BY-SA 4.0: Creative Commons Attribution-ShareAlike 4.0 International
CMWO: Commissie voor Mensgebonden Wetenschappelijk Onderzoek
EDC: electronic data capture
ICF: informed consent form
MDC: mobile data collection
mHealth: mobile health
MRS: medical record system
ODK: Open Data Kit
RDT: rapid diagnostic test
REDCap: Research Electronic Data Capture
XLS: Microsoft Excel spreadsheet
XLSForm: form standard used for the authoring of forms in a spreadsheet

Edited by A Mavragani; submitted 22.04.21; peer-reviewed by A Göritz, MDG Pimentel, I Mircheva, KM Kuo; comments to author 17.10.21; revised version received 09.12.21; accepted 06.01.22; published 16.06.22.

Please cite as:
Lambert Y, Galindo M, Sua´rez-Mutis M, Mutricy L, Sanna A, Garancher L, Cairo H, Hiwat H, Bordalo Miller J, Gomes JH, Marchesini P, Adenis A, Nacher M, Vreden S, Douine M
Tailoring Mobile Data Collection for Intervention Research in a Challenging Context: Development and Implementation in the Malakit Study
JMIR Form Res 2022;6(6):e29856
URL: https://formative.jmir.org/2022/6/e29856
doi:10.2196/29856
PMID:35708763

Original Paper

Design and Preliminary Findings of Adherence to the Self-Testing for Our Protection From COVID-19 (STOP COVID-19) Risk-Based Testing Protocol: Prospective Digital Study

Carly Herbert[1], BA; Vik Kheterpal[2], MD; Thejas Suvarna[2], BBA, BS; John Broach[3], MD, MPH, MBA; Juan Luis Marquez[4], MD, MPH; Ben Gerber[5], MD, MPH; Summer Schrader[2], BA; Christopher Nowak[2], BA; Emma Harman[2], MPH; William Heetderks[6], MD, PhD; Nisha Fahey[7], DO, ScM; Elizabeth Orvek[5], MS; Peter Lazar[5], BS; Julia Ferranto[1], BS; Kamran Noorishirazi[1], BA; Shivakumar Valpady[1], MS; Qiming Shi[1,5,8], MS; Honghuang Lin[1,9], PhD; Kathryn Marvel[1], MBA; Laura Gibson[10], MD; Bruce Barton[5], PhD; Stephenie Lemon[5], PhD; Nathaniel Hafer[8], PhD; David McManus[1,5,11], MD, ScM; Apurv Soni[1,5,9], MD, PhD

[1]Program in Digital Medicine, Department of Medicine, University of Massachusetts Chan Medical School, Worcester, MA, United States

[2]CareEvolution, Inc, Ann Arbor, MI, United States

[3]Department of Emergency Medicine, University of Massachusetts Chan Medical School, Worcester, MA, United States

[4]Washtenaw County Health Department, Washtenaw, MI, United States

[5]Department of Population and Quantitative Health Sciences, University of Massachusetts Chan Medical School, Worcester, MA, United States

[6]National Institute of Biomedical Imaging and Bioengineering, National Institutes of Health, Kelly Services, Bethesda, MD, United States

[7]Department of Pediatrics, University of Massachusetts Chan Medical School, Worcester, MA, United States

[8]University of Massachusetts Center for Clinical and Translational Science, University of Massachusetts Chan Medical School, Worcester, MA, United States

[9]Division of Clinical Informatics, Department of Medicine, University of Massachusetts Chan Medical School, Worcester, MA, United States

[10]Division of Infectious Disease, Department of Medicine, University of Massachusetts Chan Medical School, Worcester, MA, United States

[11]Division of Cardiology, Department of Medicine, University of Massachusetts Chan Medical School, Worcester, MA, United States

Corresponding Author:
Apurv Soni, MD, PhD
Division of Clinical Informatics
Department of Medicine
University of Massachusetts Chan Medical School
55 Lake Avenue North
Worcester, MA, 01655
United States
Phone: 1 774 442 4107
Email: apurv.soni@umassmed.edu

Abstract

Background: Serial testing for SARS-CoV-2 is recommended to reduce spread of the virus; however, little is known about adherence to recommended testing schedules and reporting practices to health departments.

Objective: The Self-Testing for Our Protection from COVID-19 (STOP COVID-19) study aims to examine adherence to a risk-based COVID-19 testing strategy using rapid antigen tests and reporting of test results to health departments.

Methods: STOP COVID-19 is a 12-week digital study, facilitated using a smartphone app for testing assistance and reporting. We are recruiting 20,000 participants throughout the United States. Participants are stratified into high- and low-risk groups based on history of COVID-19 infection and vaccination status. High-risk participants are instructed to perform twice-weekly testing for COVID-19 using rapid antigen tests, while low-risk participants test only in the case of symptoms or exposure to COVID-19. All participants complete COVID-19 surveillance surveys, and rapid antigen results are recorded within the smartphone app. Primary outcomes include participant adherence to a risk-based serial testing protocol and percentage of rapid tests reported to health departments.

Results: As of February 2022, 3496 participants have enrolled, including 1083 high-risk participants. Out of 13,730 tests completed, participants have reported 13,480 (98.18%, 95% CI 97.9%-98.4%) results to state public health departments with full personal identifying information or anonymously. Among 622 high-risk participants who finished the study period, 35.9% showed high adherence to the study testing protocol. Participants with high adherence reported a higher percentage of test results to the state health department with full identifying information than those in the moderate- or low-adherence groups (high: 71.7%, 95% CI 70.3%-73.1%; moderate: 68.3%, 95% CI 66.0%-70.5%; low: 63.1%, 59.5%-66.6%).

Conclusions: Preliminary results from the STOP COVID-19 study provide important insights into rapid antigen test reporting and usage, and can thus inform the use of rapid testing interventions for COVID-19 surveillance.

(JMIR Form Res 2022;6(6):e38113) doi:10.2196/38113

KEYWORDS

COVID-19; rapid antigen tests; COVID-19 testing; infectious disease; disease spread; prevention; coronavirus; adherence; reporting; mHealth; health application; mobile health; digital health; public health; surveillance; health care; smartphone app; vaccination; digital surveillance

Introduction

Despite relatively widespread vaccination for SARS-CoV-2 throughout the United States, in late February 2022, nearly 90,000 new cases of COVID-19 were reported daily in the United States among both unvaccinated and vaccinated individuals, and herd immunity remains uncertain [1]. With relaxation of masking and social distancing requirements, and many US residents returning to in-person work and schooling, widespread, accessible testing for COVID-19 is an integral component of the federal strategy to safely establish a "new normal" [2-4].

Antigen detection rapid diagnostic tests (Ag-RDTs) for COVID-19 pose great opportunity for community surveillance owing to their relative ease of use and quick turn-around time for results, making them amenable to testing outside traditional clinical environments [5]. Serial testing 2-3 times per week with Ag-RDTs is recommended to detect SARS-CoV-2 infections, specifically asymptomatic infections that comprise over 50% of total infections [6,7]. Despite the availability of this effective testing approach, little is known about adherence to this schedule outside of a controlled trial environment [8,9]. Additionally, it is unknown how COVID-19 testing strategies and schedules should be optimized based on risk factors for infection, including vaccination status [4,10,11]. At-home Ag-RDTs for COVID-19 may also challenge public health surveillance efforts owing to their reliance on individual users to carry out the tests appropriately and report their test results. Indeed, when Ag-RDTs were streamlined through the Food and Drug Administration authorization process in spring 2021, this resulted in device launches without systematic reporting mechanisms in place, leaving large potential gaps in COVID-19 surveillance data [12,13]. Public health reporting practices at the individual and practitioner levels are unknown, although likely highly varied, which present challenges in interpreting current public health data.

To fill these knowledge gaps, we are performing a longitudinal study to examine adherence to a risk-based testing protocol supported by a digital infrastructure to allow recruitment and study engagement nationwide. The goals of this prospective, site-less digital study are to leverage strong partnerships with community organizations and local health departments to assess adherence to a risk-based SARS-CoV-2 testing strategy using over-the-counter tests, and to describe participants' behavior for reporting test results to public health departments and factors associated with test reporting behavior.

Methods

Study Population and Recruitment

We are recruiting up to 20,000 participants throughout the United States who meet our predefined inclusion/exclusion criteria (Textbox 1). Study enrollment is taking place in two phases: phase 1 enrollment is restricted to Michigan residents using a convenience sample, leveraging momentum and community connections from previous COVID-19 interventions; phase 2 enrollment is open to participants anywhere in the continental United States. Recruitment efforts are being spearheaded by the RADx Community Health Equity and Engagement Team [14], with the goal of recruiting a geographic, racially, and ethnically diverse sample across the United States. Phase 2 recruitment includes respondent-driven sampling, stratified sampling, and use of digital access codes to improve representation of diverse populations in our cohort. The research team also recruits through targeted social media outreach, word of mouth, community networks, and direct communications with the support of community partners. Community partners, including community organizations and local and state health departments, are identified through the professional networks of the study team members. These partners are intentionally selected based upon their ability to serve and reach large numbers of individuals who are diverse with respect to socioeconomic status and race/ethnicity. Representatives of these organizations are sent an email description of the study and asked to distribute it through their listservs. These emails also include flyers that can be posted in community locations. The recruitment strategy is adjusted throughout the study to identify populations from regions where there is an outbreak of COVID-19 or where vaccination rates are relatively low.

The Self-Testing for Our Protection from COVID-19 (STOP COVID-19) study utilizes the Mstudy app, a custom smartphone app created within the MyDataHelps interface. The app is used to electronically collect survey data from all participants, as well as guide participants through rapid antigen testing and

interpreting their results. Residents of the mainland United States with the MyDataHelps app receive an in-app notification inviting them to participate in the study. Community partners distribute emails and flyers to interested participants with instructions to download the Mstudy app and enter a join code for the study; participants are also able to autonomously sign up through the study website. The Mstudy app is free of charge and compatible with both Apple and Android smartphone devices. All participants consent and enroll in the study using the Mstudy app. Participants less than 18 years old are required to assent, as well as receive written consent from their parents/guardians. On enrollment, participants' street address is also collected and verified against the United States Postal Service database through the digital platform to ensure that study staff can ship required testing materials directly to participants' homes. Participants are eligible for a US $50 gift card on two separate occasions throughout the study, based on their completion of surveys and rapid antigen tests.

Textbox 1. Inclusion and exclusion criteria of the STOP COVID-19 study.

Inclusion criteria

- ≥8 years of age
- Access to a smartphone
- Speak English or Spanish
- Able to provide informed consent or assent with parental consent (for participants under 18 years old)

Exclusion criteria

- Current incarceration
- Lack of mailing address
- Lack smartphone internet access
- Living outside mainland United States

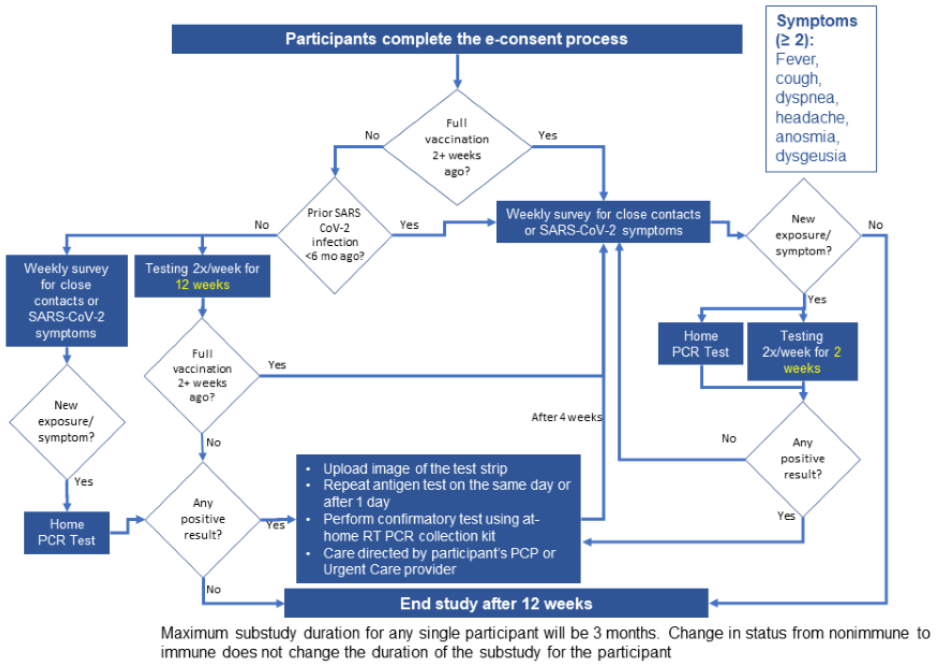
Wearable Data Collection

After enrollment, participants are asked if they would like to securely and confidentially share information from their wearable activity-tracker device (ie Fitbit, Apple Watch, and Google Fit). If participants consent to sharing wearable data, data from their activity tracker are passively collected for the 3-month study period, with no additional requirements from the participant. Sharing wearable data is optional, and participants may decline to share wearable data and still participate fully in the study. The study platform is enabled to collect measures of physical activity (eg, daily steps, distance walking or running, stairs climbed, standing time), mobility (eg, walking speed, walking asymmetry percentage, step length), vitals (eg, heart rate, body temperature, respiratory rate, resting heart rate, heart rate variability, and oxygen saturation), and sleep analysis, depending on the type of wearable device activated.

Risk Stratification

On enrollment in the study, participants are stratified into high- or low-risk groups based on SARS-CoV-2 vaccination and infection history (Figure 1, Multimedia Appendix 1). High-risk participants are defined as those who are not fully vaccinated for SARS-CoV-2, as defined by Centers for Disease Control and Prevention (CDC) guidelines, and have not had an infection in the past 6 months [15]. Low-risk participants are those who have been fully vaccinated for SARS-CoV-2 and/or infected with COVID-19 in the past 6 months.

Figure 1. STOP COVID-19 risk stratification process and testing schedule. mo: months; PCP: primary care provider; PCR: polymerase chain reaction; RT PCR: real-time polymerase chain reaction.



High-Risk Participant Study Procedures

High-risk participants are asked to test for COVID-19 twice per week for 12 consecutive weeks using an Ag-RDT (Quidel QuickVue) with the Mstudy app for testing assistance. High-risk participants receive a weekly surveillance questionnaire to assess if they are experiencing two or more new COVID-19–related symptoms (eg, fever or chills, shortness of breath, cough, loss of taste, and headache) or have had close contact with someone who tested positive for SARS-CoV-2 in the past 7 days. In the case of two or more symptoms or exposure, participants are sent a home polymerase chain reaction (PCR) collection kit via Quest Diagnostics and told to continue using the Ag-RDT at home, as the PCR test is the current diagnostic gold standard. The home PCR kit should be utilized immediately upon receipt to collect a nasal specimen and returned to Quest in the preaddressed envelope within 24 hours. Participants receive all testing and survey reminders through the app.

Low-Risk Participant Study Procedures

Low-risk participants are asked to complete weekly surveillance questionnaires to monitor for COVID-19–related symptoms or exposure to SARS-CoV-2. If a low-risk participant reports two or more COVID-19–related symptoms and/or a close exposure to someone with SARS-CoV-2, they are sent Ag-RDT kits within 48 hours and advised to test twice weekly for 2 weeks following the symptom or exposure. Additionally, participants are sent a home PCR collection kit and asked to self-collect a nasal specimen and ship it back to Quest Laboratories using the preaddressed envelope within 24 hours. After completing the 2-week testing period, low-risk participants resume completing weekly surveillance questionnaires.

Testing and Reporting Results

The Mstudy app contains detailed instructions for using the rapid tests for COVID-19 (Multimedia Appendix 1). Participants are instructed on how to correctly swab the nasal cavity and utilize the testing equipment. The app includes a timer that alerts the participant when the test is ready to be read and walks the participant through how to interpret the test results (ie, positive, negative, or inconclusive). Within the Mstudy app, participants are asked to provide an interpretation of their test result and to upload a picture of the test into the app. All positive tests are confirmed by study coordinators. If participants test positive for SARS-CoV-2, they are contacted by a physician associated with the study and advised to follow CDC guidelines for self-isolation, as well as to seek care from their primary care provider if needed. Each time a participant records their test results, they are asked if they would like to opt-in for automated reporting, either in a full or deidentified manner. Results of tests from users who agree to report either fully or anonymously are sent to a federal system (“Report Stream”) through the study mobile app (Multimedia Appendix 2, Figure S1). Results of tests from users who agree to report full (identified) information will also be sent to their respective state health department. The app is enabled to report results to all state departments of health. All reporting is done through the Mstudy app, with no additional user burden. Participants can adjust their preference for reporting test results at any time point using the app.

Questionnaire Schedule

Study participants complete weekly surveillance questionnaires through the Mstudy app, as well as additional surveys during enrollment, at baseline, after each at-home test, and on conclusion of the 3-month study period (Multimedia Appendix 2, Table S1). In addition to testing information, weekly surveillance, and information used to determine risk stratification (prior infection and vaccination), surveys gather data on participants’ demographic characteristics, COVID-19 beliefs and risk perceptions, health care utilization, medical history and health status, and reporting attitudes and perceptions. Each survey takes 2-15 minutes to complete. Participants receive

reminders to complete assigned surveys through the Mstudy app.

Data Management

All testing and questionnaire data are securely stored within rkStudio, the management platform of the Mstudy app. All collected data are deidentified using participant IDs prior to analysis and stored in the secure University of Massachusetts Chan Medical School server. PCR test results are connected to Mstudy data by participant ID.

Ethics Approval

This study protocol was approved by the Institutional Review Board of the University of Massachusetts Chan Medical School and externally by the Western Institutional Review Board-Copernicus Group (now named WCG; the Institutional Review Board number is 20213392.

Analytical Plan

Primary outcome variables of the study include adherence to the risk-based SARS-CoV-2 testing strategy, and test result reporting behaviors and motivations to participants' respective state department of health. In prespecified analyses, we will identify patterns of testing adherence over time, and identify factors associated with decreased testing adherence. We will also perform an ordinal logistic regression to assess whether participants' behavior for reporting test results to the department of health is associated with test result, vaccination status, infection history, or sociodemographic and psychosocial variables. Lastly, incidence rates of COVID-19 for participants under surveillance will be calculated overall during the study period and separately as person-time during the testing period with stratification according to participants' risk status. Risk of infection among low-risk participants will be estimated in relation to time since vaccination or infection using time-to-event analyses.

In the preliminary analysis presented, demographics from phase 1 of the study were tabulated by risk category. Testing adherence was measured among high-risk participants who finished the 12-week testing period. Adherence to the testing schedule was determined on a weekly basis, and participants were considered "adherent" to the schedule if they tested two times each week. Adherence was categorized into no, low, moderate, and high adherence groupings. The "no adherence" group included individuals who were never adherent to the testing schedule during the study period. The "low adherence" group included individuals who were adherent for 1 to 4 weeks of the 12-week study period. "Moderate adherence" included those who were adherent for 5 to 8 weeks of the study period, and "high adherence" was defined as testing twice weekly for 9 or more weeks of the study. Reporting choices were calculated at the participant level to avoid overrepresenting the reporting choices of frequent testers. Reporting choices were tabulated by risk category, test result, and adherence category, and 95% CIs were calculated using the Clopper-Pearson interval.

Results

Enrollment

The STOP COVID-19 study began enrolling phase-1 participants in August 2021, and 3496 participants have enrolled in the study to date, including 1083 high-risk participants (Figure 2 and Figure S2 in Multimedia Appendix 2). Phase 2 of recruitment started in February 2022 and is ongoing. Women make up the majority of both low- and high-risk participants, and approximately 20% of high-risk participants are children under 18 years old (Table 1). Among the adult participants, approximately 80% have a bachelor's degree or higher. Most participants are White, with Asian participants making up 8.8% and 5.0% of low- and high-risk participants, respectively (Table 1). Only approximately 1% of participants are over the age of 75 years. High-risk participants with symptoms and those with both symptoms and known exposure have a 3.66- and 1.39-times higher positivity rate for COVID-19 than low-risk participants, respectively.

Figure 2. STOP COVID-19 enrollment, fall 2021.

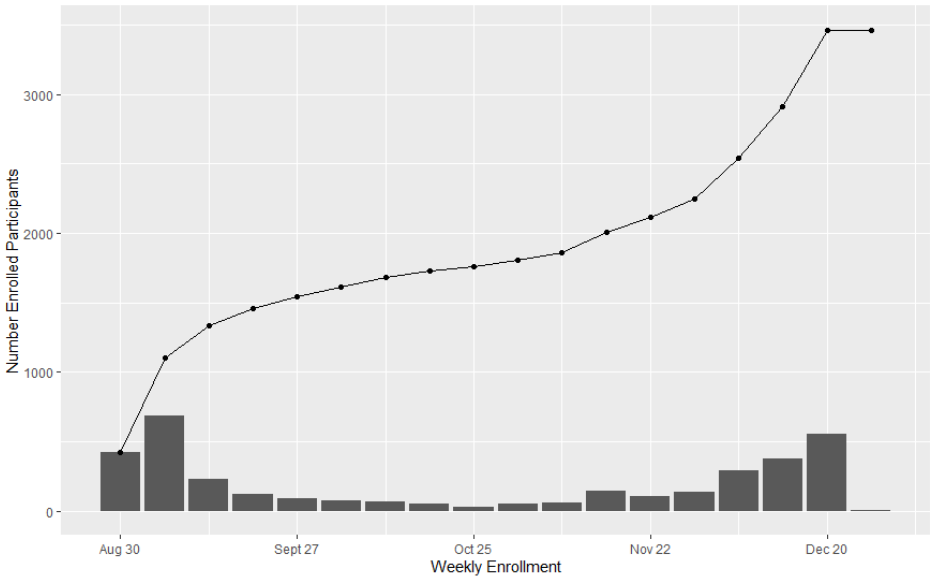


Table 1. Demographic characteristics of STOP COVID-19 participants.

Characteristics	High-risk participants (n=1083)	Low-risk participants (n=2413)
Age (years), n (%)		
8-17	202 (18.65)	135 (5.59)
18-30	106 (9.79)	423 (17.53)
31-45	375 (34.63)	1072 (44.42)
46-60	214 (19.76)	500 (20.72)
61-75	145 (13.39)	237 (9.82)
>75	19 (1.75)	22 (0.91)
Missing	22 (2.03)	24 (0.95)
Gender, n (%)		
Man	314 (28.99)	596 (24.70)
Woman	660 (60.94)	1564 (64.82)
Transgender	3 (0.28)	9 (0.37)
Nonbinary	13 (1.20)	46 (1.90)
Missing	93 (8.59)	198 (8.21)
Race, n (%)		
White	864 (79.78)	1819 (75.38)
Asian	54 (4.99)	212 (8.79)
Black/African-American	22 (2.03)	50 (2.07)
Native American/Alaskan Native	0 (0)	1 (0.04)
Native Hawaiian or Pacific Islander	0 (0)	2 (0.08)
Multiracial	32 (2.95)	71 (2.94)
Other	9 (0.83)	36 (1.49)
Missing	102 (9.42)	222 (9.20)
Hispanic, n (%)		
Yes	52 (4.80)	93 (3.85)
No	936 (86.43)	2108 (87.36)
Missing	95 (8.77)	212 (8.79)
Education level, n (%)		
Bachelor’s degree or higher	654 (60.39)	1750 (72.52)
Some college	145 (13.39)	265 (10.98)
High school graduate	33 (3.05)	80 (3.32)
Did not finish high school	134 (12.37)	110 (4.56)
Missing	117 (10.80)	208 (8.62)
Employment status, n (%)		
Working now, permanently	584 (53.92)	1398 (57.94)
Working now, temporarily	29 (2.68)	107 (4.43)
Student	190 (17.54)	332 (13.76)
Retired	96 (8.86)	146 (6.05)
Keeping house	50 (4.62)	126 (5.22)
Unemployed	21 (1.94)	77 (3.19)
Missing	113 (10.43)	227 (9.41)
Positivity of Ag-RDT[a] tests, % (95% CI)		

Characteristics	High-risk participants (n=1083)	Low-risk participants (n=2413)
Total	7.5 (5.9-9.4)	11.0 (8.8-13.5)
Symptomatic	14.3 (8.0-22.8)	3.9 (1.7-7.5)
Close contact exposure	5.2 (2.7-8.9)	6.9 (4.3-10.5)
Both symptoms and exposure	31.1 (20.8-42.9)	22.3 (17.1-28.2)
Neither	3.9 (2.4-5.9)	N/A[b]
Test reporting decisions, % (95% CI)[c]		
Full reporting	67.4 (64.4-70.3)	66.9 (63.5-70.2)
Anonymous reporting	28.2 (25.5-31.1)	30.1 (26.9-33.4)
No reporting	4.3 (3.2-5.8)	3.0 (1.9-4.4)

[a]Ag-RDT: antigen rapid diagnostic test.

[b]N/A: not applicable; low-risk participants were only eligible for testing after reporting symptoms or close contact exposure.

[c]This sample only included participants who recorded at least one test result in the Mstudy app: n=1016 for the high-risk group and n=801 for the low-risk group.

Test Reporting

As of February 2022, participants have completed over 13,730 rapid tests and have reported 13,480 (98.2%, 95% CI 97.9%-98.4%) results to their respective state public health departments. Reporting differed by test result (positive, negative, or invalid), with 3.7% (95% CI 2.0%-6.1%) of positive tests unreported in comparison to 1.8% (95% CI 1.5%-2.0%) of negative tests reported. Reporting choices did not differ significantly by risk category (Table 1). Among those who have chosen not to report their test results, the most cited reason is not wanting to be contacted by the government. Other reasons include not trusting the government, not knowing how to report, believing reporting is not useful, and being worried about missing work.

Testing Adherence and Impact on Reporting

Twelve weeks of adherence to twice-weekly serial testing was assessed among 622 high-risk participants who completed the full study period. Of these participants, 223 (35.9%) were highly adherent to the testing protocol (Table 2). The percentage of tests reported with full personal identifiers to the state department of health was significantly higher among those with high adherence, as compared to moderate and low adherent participants (Table 2). Nonreporting was significantly higher among participants with moderate adherence in comparison to those with high adherence; however, nonreporting did not differ between participants with low and high adherence (Table 2).

Table 2. Reporting decisions per test by adherence group among high-risk participants (N=622).

Variable	No adherence	Low adherence	Moderate adherence	High adherence	Total
Participants, n (%)	103 (16.6)	162 (26.0)	134 (21.5)	223 (35.9)	622 (100.0)
Total tests completed, n	13	729	1693	4145	6580
Reporting, % (95% CI)					
Full reporting	53.9 (25.1-80.8)	63.1 (59.5-66.6)	68.3 (66.0-70.5)	71.7 (70.3-73.1)	69.8 (68.7-70.9)
Anonymous reporting	30.8 (9.1-61.4)	34.8 (31.4-38.4)	28.1 (26.0-30.3)	26.5 (25.2-27.9)	27.9 (26.8-29.0)
No reporting	15.4 (1.9-45.4)	2.1 (1.2-3.4)	3.6 (2.8-4.6)	1.8 (1.4-2.2)	2.3 (2.0-2.7)

Discussion

Principal Findings

The STOP COVID-19 study is a novel longitudinal, digital study aimed at understanding adherence and public health reporting of rapid antigen testing throughout the United States, using a risk-based testing protocol. Here, we describe the study methodology, which utilizes the Mstudy app for data collection and study coordination purposes, and preliminary results from phase 1 of participant recruitment in Michigan. Phase 1 of enrollment began in August 2021, and 3496 participants in Michigan have enrolled to date, including more than 1000 high-risk participants. Digital site-less studies have many advantages in the age of COVID-19, especially to facilitate study recruitment and retention. The site-less study approach has allowed us to dynamically change the recruitment strategy throughout the study to prioritize communities with high prevalence of SARS-CoV-2, seasonal surges, or low vaccination rates in order to optimally sample communities with a high burden of COVID-19. Further, digital studies require less active study coordination than traditional site-based studies because of the ability of technology to facilitate certain tasks (ie, consenting and providing instructions for study-related activities), which allowed us to implement a risk-based testing algorithm in participants' homes nationwide while conserving

the time and resources of study personnel. The data collected from this study will offer tremendous insight into COVID-19 at-home testing behaviors, and using the digital approach was highly effective.

Adherence and Motivations for Serial Rapid Testing

There has been significant interest in rapid antigen testing on a national level as a cost-effective solution to expand serial testing for COVID-19 for surveillance and asymptomatic case detection. However, little is known about perceptions of serial testing and how individuals use rapid test results in altering their COVID-19 risk behaviors. Among the study participants engaged in rapid serial testing, the COVID-19 positivity rate was the highest among high-risk, exposed, symptomatic participants, consistent with the published literature, and supporting our approach to risk stratification [16]. Based on our preliminary data, only 35.9% of high-risk participants displayed high adherence to the weekly serial testing schedule during the study period. Although qualitative studies have identified motives for frequent testing, including a fear of unknowingly spreading COVID-19 to others, this is the first study to quantitatively evaluate adherence to serial testing programs [17]. Further, no previous studies have used a longitudinal risk-based approach to testing, which has been suggested as a key strategy in establishing a "new normal" in a world burdened by COVID-19 [11].

Perceptions of Public Health Reporting

Throughout the COVID-19 pandemic, perceptions and trust of government have strongly influenced the adoption or rejection of public health initiatives, including masking, social distancing, and vaccination. Our preliminary results showed that 14,000 rapid tests have been completed by study participants, with a full or anonymized reporting rate to participants' respective state department of public health of over 98%. We observed statistically significant differences in reporting status based on test result, with negative results being reported more than positive results.

In a study of 1420 Australian adults, individuals with higher trust in government had 6-times greater odds of adopting recommended COVID-19 avoidance behaviors, including social distancing, self-quarantine, and hand washing, than individuals with low governmental trust [18]. Additionally, high public trust in the government has been found to favorably impact the use of COVID-19 control measures, as well as increase the likelihood of vaccination for SARS-CoV-2 [19-22]. Government trust has also been highly associated with an individual's demographic characteristics and social network. A study from 178 countries found that public trust in the government during COVID-19 was positively associated with older age and good health, and negatively associated with higher education [23]. Further, we found that highly adherent participants reported a higher proportion of results to their state department of health with full identifying information than participants with moderate or low adherence to the STOP COVID-19 testing schedule. This supports the notion that COVID-19 protective (preventative) behaviors are clustered among certain individuals [24]. These results will help to shape public health messaging and initiatives, especially as rapid at-home testing becomes more accessible and widespread.

Site-Less Study Implementation Challenges

The site-less digital study approach is still relatively new in the fields of medical and public health research, although it has garnered significant interest owing to the potential benefits to participant accessibility, engagement, longitudinal data volume, and study administrative costs [25]. As the pandemic changes in severity by location and time, we believe the site-less study design is uniquely suited to provide insight into numerous communities to observe how testing behaviors change over time and vary throughout the country. In addition to fully consenting participants and collecting data through a mobile app, this study ships participants' tests for COVID-19 directly to their homes in a continuous fashion over 3 months, based on their risk category, exposures, and symptoms, which has resulted in notable implementation challenges. Although the process for rapid test distribution has been streamlined within the Mstudy app, with kit orders placed immediately based on survey responses, the PCR ordering process through Quest has been more complex. Initially, when a participant reported COVID-19 symptoms or exposure and qualified for PCR testing, they were prompted within the app to follow a link to the Quest website to register for a PCR test kit. This process was confusing and time-consuming for participants, and study coordinators were tasked with calling all eligible participants to explain the process. Further, participant registration often resulted in key identifiers being omitted from test orders and inability to match PCR test results to data from the Mstudy app. In the past 2 months, our team has worked closely with Quest to implement a roster system, by which study coordinators order PCR kits on behalf of eligible participants. This has taken significant burden off the participants and has resulted in higher participant PCR testing adherence (data not shown). As we begin phase 2 enrollment, we will continue discussions to optimize test distribution to participants, as well as other issues that arise.

Additionally, during phase 1 enrollment, convenience sampling resulted in a sample of predominately highly educated, White, female participants. Further, only 1% of the phase 1 cohort is over 75 years old, despite older adults facing the most devastating outcomes due to SARS-CoV-2 [26]. We recognize it is especially important to understand the usage of SARS-CoV-2 diagnostics and mobile health tools among older adults. During phase 2 enrollment, we will be adjusting our recruitment and sampling approaches, as detailed in the Methods section, with the goal of diversifying the cohort in terms of age, race, gender, education, and geographic location.

Study Strengths and Limitations

This is the first study to examine the reporting of COVID-19 results from Ag-RDTs and adherence to serial testing schedules without supervision, providing important data to guide the administration and use of Ag-RDTs in a real-world setting. The site-less study design offers great flexibility to recruit based on fluctuations in the pandemic, as well as to obtain a geographically diverse participant pool. Study strengths also include the use of the MyDataHelps app as a tool for reporting test results, as well as a risk-based approach to serial testing. Nevertheless, certain inclusion criteria such as the use of a smartphone may limit the generalizability of this study.

Participation, nonresponse, and attrition biases are also of concern, as individuals choosing to enroll in and comply with a longitudinal study to test for COVID-19 may differ from nonparticipants in relation to COVID-19 perceptions, education, race, and health literacy.

Conclusions

This report describes the study design and preliminary results of a longitudinal study aimed to understand individuals' testing and reporting decisions. Site-less study designs are increasingly valuable in the era of COVID-19. The data collected in this study will provide important insights into how individuals navigate their testing decisions during the COVID-19 pandemic, as well as how we should understand the data reported from rapid antigen tests for COVID-19.

Acknowledgments

This study was funded by the National Institutes of Health RADx-Tech program under 3U54HL143541-02S2 and National Institutes of Health CTSA grant UL1TR001453.

Conflicts of Interest

VK is principal, and TS, SS, CN, and EH are employees of the health care technology company CareEvolution. LG is on a scientific advisory board for Moderna for projects unrelated to COVID-19. DDM reports consulting and research support from Bristol-Myers Squibb, Pfizer, Fitbit, Flexcon, Boehringer Ingelheim, and Avania. The other authors have no conflicts of interest to report.

Multimedia Appendix 1
Screenshots of Mstudy app enrollment and testing processes.
[PPTX File , 2280 KB - formative_v6i6e38113_app1.pptx]

Multimedia Appendix 2
Supplemental tables and figures.
[DOCX File , 126 KB - formative_v6i6e38113_app2.docx]

References

1. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020 May;20(5):533-534 [FREE Full text] [doi: 10.1016/S1473-3099(20)30120-1] [Medline: 32087114]
2. Gniazdowski V, Paul Morris C, Wohl S, Mehoke T, Ramakrishnan S, Thielen P, et al. Repeated coronavirus disease 2019 molecular testing: correlation of Severe Acute Respiratory Syndrome Coronavirus 2 culture with molecular assays and cycle thresholds. Clin Infect Dis 2021 Aug 16;73(4):e860-e869 [FREE Full text] [doi: 10.1093/cid/ciaa1616] [Medline: 33104776]
3. Paltiel AD, Zheng A, Walensky RP. Assessment of SARS-CoV-2 screening strategies to permit the safe reopening of college campuses in the United States. JAMA Netw Open 2020 Jul 01;3(7):e2016818 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.16818] [Medline: 32735339]
4. Christie A, Brooks JT, Hicks LA, Sauber-Schatz EK, Yoder JS, Honein MA, CDC COVID-19 Response Team. Guidance for implementing COVID-19 prevention strategies in the context of varying community transmission levels and vaccination coverage. MMWR Morb Mortal Wkly Rep 2021 Jul 27;70(30):1044-1047. [doi: 10.15585/mmwr.mm7030e2] [Medline: 34324480]
5. Guidance for Antigen Testing for SARS-CoV-2 for Healthcare Providers Testing Individuals in the Community. Centers for Disease Control and Prevention. 2021. URL: https://www.cdc.gov/coronavirus/2019-ncov/lab/resources/antigen-tests-guidelines.html [accessed 2022-06-05]
6. Moghadas SM, Fitzpatrick MC, Sah P, Pandey A, Shoukat A, Singer BH, et al. The implications of silent transmission for the control of COVID-19 outbreaks. Proc Natl Acad Sci U S A 2020 Jul 28;117(30):17513-17515 [FREE Full text] [doi: 10.1073/pnas.2008373117] [Medline: 32632012]
7. Johansson MA, Quandelacy TM, Kada S, Prasad PV, Steele M, Brooks JT, et al. SARS-CoV-2 transmission from people without COVID-19 symptoms. JAMA Netw Open 2021 Jan 04;4(1):e2035057 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.35057] [Medline: 33410879]
8. Smith R, Gibson L, Martinez P, Ke R, Mirza A, Conte M, et al. Longitudinal assessment of diagnostic test performance over the course of acute SARS-CoV-2 infection. J Infect Dis 2021 Sep 17;224(6):976-982. [doi: 10.1093/infdis/jiab337] [Medline: 34191025]
9. Abbasi J. Antigen testing every 3 days is highly sensitive for SARS-CoV-2. JAMA 2021 Aug 10;326(6):473. [doi: 10.1001/jama.2021.12620] [Medline: 34374735]

10. Mina MJ, Andersen KG. COVID-19 testing: one size does not fit all. Science 2021 Jan 08;371(6525):126-127. [doi: 10.1126/science.abe9187] [Medline: 33414210]
11. Emanuel EJ, Osterholm M, Gounder CR. A national strategy for the "new normal" of life with COVID. JAMA 2022 Jan 18;327(3):211-212. [doi: 10.1001/jama.2021.24282] [Medline: 34989789]
12. Court E. Many U.S. at-home COVID-19 test results could be going unreported. Time Magazine. 2021 Aug 24. URL: https://time.com/6092343/at-home-covid-tests-data/ [accessed 2022-06-05]
13. Shuren J, Stenzel T. Coronavirus (COVID-19) Update: FDA takes steps to streamline path for COVID-19 screening tools, provides information to help groups establishing testing programs. U.S. Food and Drug Administration. 2021 Mar 16. URL: https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-takes-steps-streamline-path-covid-19-screening-tools-provides [accessed 2022-06-05]
14. Gibson L, Fahey N, Hafer N, Buchholz B, Dunlap D, Murphy R, et al. The RADx Tech Clinical Studies Core: a model for academic based clinical studies. IEEE Open J Eng Med Biol 2021;2:152-157 [FREE Full text] [doi: 10.1109/OJEMB.2021.3070830] [Medline: 34192287]
15. Wu J, Liang B, Chen C, Wang H, Fang Y, Shen S, et al. SARS-CoV-2 infection induces sustained humoral immune responses in convalescent patients following symptomatic COVID-19. Nat Commun 2021 Mar 22;12(1):1813. [doi: 10.1038/s41467-021-22034-1] [Medline: 33753738]
16. Scobie HM, Johnson AG, Suthar AB, Severson R, Alden NB, Balter S, et al. Monitoring incidence of COVID-19 cases, hospitalizations, and deaths, by vaccination status - 13 U.S. jurisdictions, April 4-July 17, 2021. MMWR Morb Mortal Wkly Rep 2021 Sep 17;70(37):1284-1290. [doi: 10.15585/mmwr.mm7037e1] [Medline: 34529637]
17. Wanat M, Logan M, Hirst JA, Vicary C, Lee JJ, Perera R, et al. Perceptions on undertaking regular asymptomatic self-testing for COVID-19 using lateral flow tests: a qualitative study of university students and staff. BMJ Open 2021 Sep 02;11(9):e053850 [FREE Full text] [doi: 10.1136/bmjopen-2021-053850] [Medline: 34475190]
18. Seale H, Heywood AE, Leask J, Sheel M, Thomas S, Durrheim DN, et al. COVID-19 is rapidly changing: examining public perceptions and behaviors in response to this evolving pandemic. PLoS One 2020 Jun 23;15(6):e0235112 [FREE Full text] [doi: 10.1371/journal.pone.0235112] [Medline: 32574184]
19. Pak A, McBryde E, Adegboye OA. Does high public trust amplify compliance with stringent COVID-19 government health guidelines? A multi-country analysis using data from 102,627 individuals. Risk Manag Healthc Policy 2021;14:293-302. [doi: 10.2147/RMHP.S278774] [Medline: 33542664]
20. Han Q, Zheng B, Cristea M, Agostini M, Bélanger JJ, Gützkow B, PsyCorona Collaboration, et al. Trust in government regarding COVID-19 and its associations with preventive health behaviour and prosocial behaviour during the pandemic: a cross-sectional and longitudinal study. Psychol Med 2021 Mar 26:1-11 [FREE Full text] [doi: 10.1017/S0033291721001306] [Medline: 33769242]
21. Latkin CA, Dayton L, Yi G, Konstantopoulos A, Boodram B. Trust in a COVID-19 vaccine in the U.S.: a social-ecological perspective. Soc Sci Med 2021 Feb;270:113684 [FREE Full text] [doi: 10.1016/j.socscimed.2021.113684] [Medline: 33485008]
22. Arvanitis M, Opsasnick L, O'Conor R, Curtis L, Vuyyuru C, Yoshino Benavente J, et al. Factors associated with COVID-19 vaccine trust and hesitancy among adults with chronic conditions. Prev Med Rep 2021 Dec;24:101484 [FREE Full text] [doi: 10.1016/j.pmedr.2021.101484] [Medline: 34306998]
23. Gozgor G. Global evidence on the determinants of public trust in governments during the COVID-19. Appl Res Qual Life 2022;17(2):559-578 [FREE Full text] [doi: 10.1007/s11482-020-09902-6] [Medline: 33564341]
24. Nishimi K, Borsari B, Marx BP, Rosen RC, Cohen BE, Woodward E, et al. Clusters of COVID-19 protective and risky behaviors and their associations with pandemic, socio-demographic, and mental health factors in the United States. Prev Med Rep 2022 Feb;25:101671 [FREE Full text] [doi: 10.1016/j.pmedr.2021.101671] [Medline: 34926133]
25. Inan OT, Tenaerts P, Prindiville SA, Reynolds HR, Dizon DS, Cooper-Arnold K, et al. Digitizing clinical trials. NPJ Digit Med 2020 Jul 31;3(1):101. [doi: 10.1038/s41746-020-0302-y] [Medline: 32821856]
26. Yanez N, Weiss N, Romand J. COVID-19 mortality risk for older men and women. BMC Public Health 2020:20-1742 [FREE Full text]

Abbreviations

Ag-RDT: antigen rapid diagnostic test
CDC: Centers for Disease Control and Prevention
PCR: polymerase chain reaction
STOP COVID-19: Self-Testing for Our Protection from COVID-19

Edited by A Mavragani; submitted 19.03.22; peer-reviewed by L Guo; comments to author 15.04.22; revised version received 11.05.22; accepted 29.05.22; published 16.06.22.

Please cite as:
Herbert C, Kheterpal V, Suvarna T, Broach J, Marquez JL, Gerber B, Schrader S, Nowak C, Harman E, Heetderks W, Fahey N, Orvek E, Lazar P, Ferranto J, Noorishirazi K, Valpady S, Shi Q, Lin H, Marvel K, Gibson L, Barton B, Lemon S, Hafer N, McManus D, Soni A
Design and Preliminary Findings of Adherence to the Self-Testing for Our Protection From COVID-19 (STOP COVID-19) Risk-Based Testing Protocol: Prospective Digital Study
JMIR Form Res 2022;6(6):e38113
URL: https://formative.jmir.org/2022/6/e38113
doi:10.2196/38113
PMID:35649180

Original Paper

Feasibility of Using Games to Improve Healthy Lifestyle Knowledge in Youth Aged 9-16 Years at Risk for Type 2 Diabetes: Pilot Randomized Controlled Trial

Ralph Maddison[1,2], PhD; Nilufar Baghaei[3,4], PhD; Amanda Calder[2], PhD; Rinki Murphy[5], MD; Varsha Parag[2], MSc; Ihirangi Heke[6], PhD; Rosie Dobson[2], PhD; Samantha Marsh[2], PhD

[1]Institute for Physical Activity and Nutrition, Deakin University, Geelong, Australia
[2]National Institute for Health Innovation, University of Auckland, Auckland, New Zealand
[3]Games and Extended Reality Lab, Massey University, Auckland, New Zealand
[4]School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, Australia
[5]Department of Medicine, Middlemore Hospital, University of Auckland, Auckland, New Zealand
[6]Heke Consulting, Auckland, New Zealand

Corresponding Author:
Ralph Maddison, PhD
Institute for Physical Activity and Nutrition
Deakin University
1 Gheringhap Street
Geelong, 3220
Australia
Phone: 61 392446218
Email: ralph.maddison@deakin.edu.au

Abstract

Background: Mobile games can be effective and motivating tools for promoting children's health.

Objective: We aimed to determine the comparative use of 2 prototype serious games for health and assess their effects on healthy lifestyle knowledge in youth aged 9-16 years at risk for type 2 diabetes (T2D).

Methods: A 3-arm parallel pilot randomized controlled trial was undertaken to determine the feasibility and preliminary effectiveness of 2 serious games. Feasibility aspects included recruitment, participant attitudes toward the games, the amount of time the participants played each game at home, and the effects of the games on healthy lifestyle and T2D knowledge. Participants were allocated to play *Diabetic Jumper* (n=7), *Ari and Friends* (n=8), or a control game (n=8). All participants completed healthy lifestyle and T2D knowledge questionnaires at baseline, immediately after game play, and 4 weeks after game play. Game attitudes and preferences were also assessed. The primary outcome was the use of the game (specifically, the number of minutes played over 4 weeks).

Results: In terms of feasibility, we were unable to recruit our target of 60 participants. In total, 23 participants were recruited. Participants generally viewed the games positively. There were no statistical differences in healthy lifestyle knowledge or diabetes knowledge over time or across games. Only 1 participant accessed the game for an extended period, playing the game for a total of 33 min over 4 weeks.

Conclusions: It was not feasible to recruit the target sample for this trial. The 2 prototype serious games were unsuccessful at sustaining long-term game play outside a clinic environment. Based on positive participant attitudes toward the games, it is possible to use these games or similar games as short-term stimuli to engage young people with healthy lifestyle and diabetes knowledge in a clinic setting; however, future research is required to explore this area.

Trial Registration: Australia New Zealand Clinical Trials Registry ACTRN12619000380190; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=377123

(JMIR Form Res 2022;6(6):e33089) doi:10.2196/33089

KEYWORDS

children's health; diabetes mellitus; type 2 diabetes; experimental games; recruitment

Introduction

As the prevalence of type 2 diabetes (T2D) in New Zealand's younger population increases, the challenge of designing effective and engaging ways and means of preventing the disease also increases [1,2]. Recently, there has been an upsurge of interest in the role of using computer and video games to enhance health outcomes for youth [3,4]. These "serious video games" are designed to entertain players, as they educate, train, or change behavior [3]. For example, games for health are serious video games focused on health [5], with most having some positive outcomes [5], such as effectively promoting dietary change among youth [6]. Indeed, existing video games have helped children between the ages of 10 and 12 years make healthier diet and physical activity choices, with children as young as 9 years understanding the games [7]. Games for health could help young people with T2D attain better self-management by making education about healthy lifestyles more engaging and fun [8]. However, there is a dearth of games designed specifically for youth with T2D [5].

One game that was developed to promote knowledge of the interaction of physical activity behaviors and diet with blood glucose, and knowledge of blood glucose monitoring in young people with T2D resulted in promising outcomes [9,10]. In a preliminary pre-post pilot trial, 12 children aged 9-13 years without T2D or prior experience playing videogames found the game to be fun. Moreover, their engagement was high, and they felt part of a creative and dynamic game community. Importantly, the game was found to enhance the children's knowledge of healthy diet and lifestyle choices [9,10]. By taking into account user preferences and abilities from this prototype game, we developed 2 new serious games for health. We undertook a pilot randomized controlled study to determine the feasibility and preliminary effectiveness of the 2 serious games. Specifically, the feasibility aspects were recruitment, participant acceptability, preferences and attitudes toward the game, and the amount of time participants played each game in their own environments.

Methods

Study Design and Participants

A 3-arm parallel pilot randomized controlled trial (RCT) was undertaken in Auckland, New Zealand between April and September 2019. Eligible participants were aged between 9 and 16 years, had a family history of T2D, were overweight or obese for their age according to the Cole International cutoff points for BMI, or had been told by their doctor that they were at risk for T2D. Eligible participants were also required to have access to an Android device, be able to provide assent (if under the age of 16 years) or consent (if 16 years or older), be able to speak and understand English, and live in the Auckland region. Eligible parents or caregivers were over the age of 18 years, could provide written informed consent on behalf of the child (if the child was under 16 years old), and could speak and understand English.

Ethics Approval

Ethics approval was obtained from the University of Auckland Human Participants Ethics Committee (ref# 022616). The Consolidated Standards of Reporting Trials (CONSORT) guidelines for reporting parallel group RCTs were followed [11] (Multimedia Appendix 1).

Recruitment

Participants were recruited from April to September 2019 through posts on websites and social media (Neighbourly, Facebook, and Twitter) by the University of Auckland and affiliates, and paid Facebook and Google advertisements that targeted Auckland parents. Specifically, we included 7 weeks of continuous Facebook advertising, 4 weeks of continuous Google advertising, and 3 to 4 separate posts on Twitter and Facebook. We also placed recruitment flyers in waiting rooms or reception areas of community centers and businesses in the health or youth sector (eg, Waitakere Foot Clinics, YMCA, Sport Auckland, and ProCare-associated general practices). Participants were also recruited through face-to-face contact by the research assistant attending 14 separate community events and talking to parents, or through a trusted third party such as a community program. After 3 months of low recruitment, our approach was updated to include schools as potential third parties. Three schools were approached, and one consented to be involved. Further, the steering committee made 3 amendments to the trial in a bid to reduce barriers to participation and meet recruitment targets. We updated the study eligibility criteria so that participants no longer required access to an Android device and were no longer restricted to living in the Auckland region. Additionally, advertising materials no longer mentioned that we were targeting overweight/obese children for this intervention [12]. These amendments were supplemented by approaching more than 30 potential third parties, including New Zealand teachers (on the Facebook page), 9 Christchurch schools, researchers from other New Zealand Universities, and Diabetes NZ and affiliated clinics.

Procedure

The parent or caregiver and child attended a 1-hour clinic-style study assessment with a trained research assistant, which was held at the National Institute for Health Innovation, Auckland or at the child's school. The child completed questions online pertaining to a healthy lifestyle and T2D (Multimedia Appendix 2) before being allocated at random to play 1 of 3 games: *Ari and Friends*, *Diabetic Jumper*, and *Doodle Jump* (control game). Participants were randomized in a 1:1:1 ratio, using a computer-generated random sequence with block randomization involving block sizes of 3 created by the study statistician. Immediately after the initial game play and 4 weeks after the meeting, the children completed the online healthy lifestyle and T2D questionnaires again, with additional questions on game attitudes and preferences (Multimedia Appendix 3). Participants were invited to play the game over the 4 weeks as much as they liked in between assessments. Game play was recorded by the game software. All participants were given two NZ $40 (US $26) supermarket vouchers: one (in person) at baseline and the other (through post) at follow-up.

Intervention

The first game *Ari and Friends* was implemented in Android and was adapted from a Mario Brothers open-source platform (Figure 1), and the second game *Diabetic Jumper* (also in Android) was adapted from a *Doodle Jump* open-source platform (Figure 2). The research team modified concepts from the respective games to suit this project. Initial development involved prototyping the game features with the research team and some tertiary students before formal evaluation was undertaken. Using user-design principles, we recruited a convenience sample of potential end users (children of the same age as the trial) who played early iterations of the respective games and provided feedback to developers and the research team in a workshop format. In total, there were 4 such iterative development workshops prior to the final pilot study, and they were based on feedback that users were happy playing the final versions of the games on the mobile platforms.

Both of the final games included features designed to increase children's knowledge of energy intake and expenditure. For example, by selecting certain foods as the player progressed through the levels, they began to understand the impact of physical activity (stamina to run or jump) and of sweet foods and drinks on their energy levels. Moreover, questions about healthy lifestyle and T2D were interspersed in the game (Multimedia Appendix 4), which could boost power levels if they answered correctly (Figure 3). Players were informed if their answers were correct, and if not, the correct answer was highlighted. For example, at level 1 of the game, children were asked, "How much moderate and/or vigorous physical activity should you do in a week?" Various options were then provided (eg, "at least half an hour a day at least 3 times a week" or "at least 1 hour a day every day of the week"), which the child could select. Children could play the game for a maximum of 15 min per session.

Figure 1. In-game image from *Ari and Friends*.

Figure 2. In-game image from *Diabetic Jumper*.

Figure 3. In-game example of the questions asked and answered in *Ari and Friends*.

Control

The control game *Doodle Jump* did not have any information on T2D or healthy lifestyle behaviors. The object of the game was similar to *Diabetic Jumper*, with the aim to use platforms to continue jumping upwards without falling. Children allocated to the control game also played for 15 min per session.

Outcomes

The primary outcome was the number of hours played over a 4-week period recorded via the game software. Secondary outcomes were assessed via questionnaires 15 min (±5 min) after game play during the baseline session and at the 4-week follow-up. Secondary outcomes included (1) change in healthy lifestyle knowledge from baseline, (2) change in diabetes-specific knowledge from baseline, and (3) attitudes and preferences assessed using open-ended and free-text questions regarding the games. The questionnaires were designed specifically for this study and contained the same healthy lifestyle and T2D questions that boosted power levels in the games. Specifically, there were 18 multiple-choice questions about healthy lifestyles (specifically about physical activity, screen time, sleep, sedentary behavior, sources of health information, and nutrition) and 9 true/false questions about T2D. Questions about specific healthy lifestyle topics featured in different levels of the 2 intervention games. For example, level 1 in both games contained questions about physical activity and diabetes, level 2 contained questions about nutrition and diabetes, and the remaining 4 levels contained other healthy lifestyle topics. See Multimedia Appendix 4 for the full list of questions.

Statistical Analysis

Our recruitment target was 60 participants, with 20 participants per study arm. As this was a pilot trial, our recruitment target was not powered to detect significant differences between groups. Study data were collected using REDCap, and all analyses were performed using SAS version 9.4 (SAS Institute). Baseline and follow-up variables were summarized according to group and descriptive summary statistics provided. The change from baseline in continuous outcomes was analyzed using ANOVA for normally distributed data and the Kruskal-Wallis test for nonnormally distributed data.

Results

Participants

In terms of feasibility, we were unable to recruit our target of 60 participants. In total, 23 participants (mean age 11 years, SD 1 year) were recruited between April and September 2019 through direct methods (Multimedia Appendix 5). Four participants were recruited via advertisements, while 19 participants were recruited from 1 Auckland school. Seven children were randomized to *Diabetic Jumper*, 8 to *Ari and Friends*, and 8 to the control game. Participants identified predominantly as Samoan (n=15) and Tongan (n=6), and their BMI ranged from 22 to 46 kg/m^2. See Table 1 for details.

Table 1. Participant demographic characteristics.

Characteristic	Intervention: *Diabetic Jumper* (n=7)	Intervention: *Ari and Friends* (n=8)	Active control (n=8)
Age (years), mean (SD)	11 (1.1)	11.3 (1.4)	11.3 (2.9)
Sex (female), n (%)	2 (30)	5 (60)	7 (90)
Weight (kg), mean (SD)	71.8 (23.3)	88.3 (29.0)	84.4 (29.2)
Height (m), mean (SD)	1.52 (0.1)	1.56 (0.1)	1.56 (0.3)
BMI (kg/m^2), mean (SD)	29.9 (7.2)	35.5 (8.7)	33.6 (6.9)
Ethnicity (Samoan), n (%)	4 (57)	4 (50)	7 (88)

Game Play

All participants played the games during the initial visit (mean duration 15 min, SD 2 min). Only 1 participant accessed the game after the testing session. The participant played *Diabetic Jumper* thrice in the first evening and twice the next day (total 33 min).

Diabetes Knowledge

The mean correct scores for diabetes knowledge ranged from 4.6 to 6.6 (51%-73%) for each game at baseline, 5.6 to 6.0 (62%-66%) at the 15-min follow-up, and 5.3 to 6.6 (58%-73%) at the 4-week follow-up (Table 2). There was no statistically significant difference in the mean change in the diabetes knowledge score at the 15-min follow-up (P=.93) or the 4-week follow-up (P=.10).

Table 2. Mean correct scores in the diabetes and healthy lifestyle knowledge questionnaires across the 2 intervention games (*Ari and Friends* and *Diabetic Jumper*) and the control game.

Outcome	Game						*P* value
	Ari and Friends		*Diabetic Jumper*		Control game		
	N	Score[a], mean (SD)	N	Score[a], mean (SD)	N	Score[a], mean (SD)	
Diabetes knowledge questionnaire (maximum score=9)							
Baseline	8	6.6 (1.4)	7	4.6 (1.0)	8	5.8 (2.6)	
15-min follow-up	8	5.6 (1.4)	7	5.6 (1.5)	8	6.0 (1.8)	
4-week follow-up	8	5.3 (0.9)	7	5.4 (2.1)	8	6.6 (1.3)	
Change (15-min − baseline)	8	−1.0 (1.7)	7	1.0 (1.7)	8	0.3 (1.7)	.93
Change (4-week − baseline)	8	−1.4 (1.5)	7	0.9 (2.7)	8	0.9 (2.5)	.10
Healthy lifestyle knowledge questionnaire (maximum score=18)							
Baseline	8	9.4 (2.7)	7	10.0 (3.1)	8	10.5 (3.1)	
15-min follow-up	8	10.4 (3.1)	7	10.6 (3.3)	8	11.4 (3.2)	
4-week follow-up	8	9.1 (2.2)	6	11.3 (1.9)	8	10.6 (3.6)	
Change (15-min − baseline)	8	1.0 (2.0)	7	0.6 (1.4)	8	0.9 (1.1)	.86
Change (4-week − baseline)	8	−0.3 (1.4)	6	0.8 (2.1)	8	0.1 (2.4)	.64

[a]Higher scores indicate a better outcome (more correct answers).

Healthy Lifestyle Behavior Knowledge

The mean correct scores for healthy lifestyle behavior knowledge ranged from 9.4 to 10.5 (52%-58%) for each game at baseline, 10.4 to 11.4 (57%-63%) at the 15-min follow-up, and 9.1 to 11.3 (50%-63%) at the 4-week follow-up (Table 2). There was no statistically significant difference in the mean change in the healthy lifestyle behavior knowledge score at the 15-min follow-up (P=.86) or the 4-week follow-up (P=.64).

Attitudes and Preferences at the 15-min Follow-up

For *Ari and Friends*, 5 of the 8 (63%) participants indicated that the controls were easy to use and that the game was fun, but only 2 of the 8 (25%) participants said that they would recommend the game to their friends.

For *Diabetic Jumper*, 6 of the 7 (85%) participants indicated that the controls were easy to use and the game was fun, and 3 of the 7 (42%) participants indicated that they would recommend the game to their friends.

Overall, 16 of the 23 (69%) participants reported that their parents allowed them to play video games in general “a bit” or “a lot,” and 13 of the 23 (56%) participants reported that their parents restricted their video game play to some extent. Moreover, 10 of the 23 (43%) participants were unsure if their parents would allow them to play the games from the study.

Discussion

Principal Findings

Overall, this pilot trial sought to determine the feasibility and preliminary knowledge effects of using 2 prototype serious games for health to improve healthy lifestyle knowledge in youth aged 9-16 years at risk for T2D. The results showed no evidence of improved healthy lifestyle or diabetes knowledge immediately after playing the games or after 4 weeks. Despite all participants playing the games for the allocated 15 min and the mostly positive feedback about each of the 2 games, only 1 participant played the game after leaving the clinic, and this participant played only during the following 2 days. These results coupled with low participant numbers indicated that the games were not engaging enough to result in sustained play and that the intervention methodology was not feasible. However, there may still be potential for using these games in a clinic setting as tools with which to engage youth in healthy lifestyle and T2D knowledge alongside individual clinical advice.

Comparison With Prior Work

The operation of the 2 prototype games and the method of data extraction from the games were successful. There were no technical issues experienced in either of the games, and extraction of data on play duration, question answers, and food choices from the game software was achieved without any issues. Moreover, the children enjoyed playing both games. In combination, these 3 outcomes indicated that the games were reliable tools to briefly engage with young people who are at risk of diabetes or with diabetes. These outcomes align with the purpose of diabetes-based games for health, making the learning of glycemic control practices engaging and fun [8]. Contrary to previous games for health research, however, there was no evidence from this pilot trial that playing either of the games led to a change in the knowledge of healthy lifestyle or diabetes in the short or long term. This outcome was unexpected as many health-related video games have demonstrated positive outcomes [6]. Indeed, some games for health have resulted in children making positive changes in their diet and physical activity choices [7]. It was difficult to determine whether the lack of effects of these games was a result of the games themselves or a lack of statistical power to detect a change in knowledge over time. Unfortunately, the trial was hampered by recruitment issues. Despite extensive efforts to recruit participants, less than 40% of the recruitment target was achieved over 7 months.

Previous studies involving games for health with hard-to-reach populations, such as youth at risk for diabetes, have ranged from case studies (n=1) to large experimental studies (n=200) [4]. Larger studies have reported using the same recruitment methods as those in this trial, for example, recruiting entire schools [13], collaborating with third parties (eg, recruiting diabetes clinics to gain access to adolescents with diabetes [14,15] and having the trial endorsed by a trusted and known staff member in Māori communities [16]), and using online communication tools and social media platforms [10,17,18]. Of note is that many of these studies did not report the initial recruitment target or the detailed recruitment methods [19,20].

We propose 2 primary reasons for why the recruitment methods in this trial were not as successful as those in previous trials involving games for health. First, there was a teacher strike [21] and a national measles outbreak during the recruitment period. As a result, running a trial may not have been attractive/practical for teachers during this time. Second, it takes time to build a trusting and respectful relationship between communities (eg, schools and health organizations) and research institutes [22]. Over-researched communities are increasingly wary of potential negative effects from having institutes (outsiders) conduct research [23]. Potential negative effects can be negated through including community members in lead roles (which has improved community self-efficacy through problem solving and making decisions) and promoting independence from the researcher by providing tools and resources to the community once the trial is complete [16,24]. Positive relations have also developed between research institutes and communities when health research projects are co-designed [25]. We believe that public health researchers would benefit from investing time and resources into building strong relationships with third parties before designing a trial. Future research should consider a long notice period when planning recruitment, establish relationships with communities and schools, and keep in mind that schools present an annual charter outlining their plans and aims up to 5 years in the future [26].

Strengths and Limitations

Key strengths of this trial include the considerable formative work to develop the games, the RCT design, and the objective collection of game play directly via the game software. One design limitation that hindered recruitment was having 2 consent processes when recruiting through a school. The school principal consented to conducting research on school grounds during or after school time with the pupils. Due to the age of the participants, parents also consented to participate or for a school or YMCA staff member to be present in lieu of the parent. Consequently, children were responsible for taking the consent forms home to their parents and returning the signed forms to a staff member. Forms were subsequently lost or forgotten and never returned. These consent processes were designed to remove the burden from parents having to attend baseline sessions and to keep their children involved. Unfortunately, the children received the burden of looking after the form. It is possible that recruitment could have been more successful with an alternative process (eg, online forms) to ease the burden on the participant dyad.

Three potential sources of bias should be noted. First, parents, caregivers, or staff members were present during the clinic-type sessions. The presence of authority figures might have influenced participants' attitudes and preferences toward the games in a bid to be polite, please the authority figure, or behave oneself. This "effort to please" could explain why attitudes and preferences were positive, yet the games were not played during follow-up. Second, the eligibility criteria stipulated that participants had to have access to an Android device, not necessarily that they had to own one themselves. Devices were predominantly owned by parents or elder siblings, and the children might not have the opportunity to use the devices at home, limiting their opportunity to play the games. Third, the research assistant read the questions aloud to many of the younger participants who were struggling to read themselves. The results showed no evidence of change in knowledge, indicating that any unintentional inflection was undetected by the limited sample size or, more likely, was indicative that the participants did not understand the questions at all and guessed.

Future Directions

The original purpose of the 2 prototype games was for children to play them while waiting with their parents to see the doctor [10]. Initial game play in a waiting room at a diabetes clinic could be a catalyst for discussions among clinicians, parents, and young people about the future risk and management of T2D, including healthy lifestyles. If a future trial is designed to mimic this situation, family members could also play the games. We propose to investigate whether these games can be used to build conversation pathways about T2D and healthy lifestyles between families and clinicians.

Conclusion

It was not feasible to recruit the target sample for this trial. Our 2 prototype serious games were unsuccessful at sustaining long-term play outside a clinic environment. Based on positive participant attitudes toward the games, it is possible to use these games or similar games as short-term stimuli to engage young people with healthy lifestyle and diabetes knowledge in a clinic setting; however, future research is required to explore this area.

Acknowledgments

The authors wish to acknowledge the Health Research Council of New Zealand for funding the study. This work was funded as part of a Health Research Council of New Zealand Feasibility Grant (17/553). We would like to thank all participants and schools for their involvement in the study. We would also like to thank Ilona Halim and John Casey for their work in developing the prototype games.

Authors' Contributions

RMaddison conceived the study and procured funding. RMaddison drafted the initial version of the paper and approved the version to be published. SM and AC provided inputs on various drafts. All investigators reviewed versions of the paper and provided inputs.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT checklist.
[PDF File (Adobe PDF File), 67 KB - formative_v6i6e33089_app1.pdf]

Multimedia Appendix 2
Type 2 diabetes and lifestyle questions.
[PDF File (Adobe PDF File), 48 KB - formative_v6i6e33089_app2.pdf]

Multimedia Appendix 3
Follow-up questionnaire.
[PDF File (Adobe PDF File), 54 KB - formative_v6i6e33089_app3.pdf]

Multimedia Appendix 4
Embedded game questions.
[PDF File (Adobe PDF File), 116 KB - formative_v6i6e33089_app4.pdf]

Multimedia Appendix 5
CONSORT recruitment diagram.
[PDF File (Adobe PDF File), 45 KB - formative_v6i6e33089_app5.pdf]

References

1. Ministry of Health. Living Well with Diabetes: A plan for people at high risk of or living with diabetes 2015–2020. Ministry of Health, New Zealand. URL: https://www.health.govt.nz/system/files/documents/publications/living-well-with-diabetes-oct15.pdf [accessed 2022-05-17]
2. Jefferies C, Carter P, Reed PW, Cutfield W, Mouat F, Hofman PL, et al. Pediatr Diabetes 2012 Jun 07;13(4):294-300. [doi: 10.1111/j.1399-5448.2011.00851.x]
3. Stokes B. Videogames have changed: time to consider 'Serious Games'? The Development Education Journal 2005;11(3):1-8 [FREE Full text]
4. Sardi L, Idri A, Fernández-Alemán JL. A systematic review of gamification in e-Health. J Biomed Inform 2017 Jul;71:31-48 [FREE Full text] [doi: 10.1016/j.jbi.2017.05.011] [Medline: 28536062]
5. Baranowski T, Buday R, Thompson DI, Baranowski J. Playing for real: video games and stories for health-related behavior change. Am J Prev Med 2008 Jan;34(1):74-82 [FREE Full text] [doi: 10.1016/j.amepre.2007.09.027] [Medline: 18083454]
6. Baranowski T, Baranowski J, Cullen KW, Marsh T, Islam N, Zakeri I, et al. Squire's Quest!. American Journal of Preventive Medicine 2003 Jan;24(1):52-61. [doi: 10.1016/s0749-3797(02)00570-6] [Medline: 12554024]

7. Baranowski T, Baranowski J, Thompson D, Buday R, Jago R, Griffith MJ, et al. Video game play, child diet, and physical activity behavior change a randomized clinical trial. Am J Prev Med 2011 Jan;40(1):33-38 [FREE Full text] [doi: 10.1016/j.amepre.2010.09.029] [Medline: 21146765]
8. Chen G, Baghaei N, Sarrafzadeh A, Manford C, Marshall S, Court G. Designing games to educate diabetic children. In: OzCHI '11: Proceedings of the 23rd Australian Computer-Human Interaction Conference. 2011 Presented at: 23rd Australian Computer-Human Interaction Conference; November 28, 2011-December 2, 2011; Canberra, Australia p. 72-75. [doi: 10.1145/2071536.2071546]
9. Baghaei N, Nandigam D, Casey J, Direito A, Maddison R. Evaluating mobile games for diabetes education. In: Proceedings of the 23rd International Conference on Computers in Education, ICCE 2015. 2015 Presented at: 23rd International Conference on Computers in Education; November 30, 2015-December 4, 2015; Hangzhou, China p. 521-526.
10. Baghaei N, Nandigam D, Casey J, Direito A, Maddison R. Diabetic Mario: Designing and Evaluating Mobile Games for Diabetes Education. Games Health J 2016 Aug;5(4):270-278. [doi: 10.1089/g4h.2015.0038] [Medline: 27304882]
11. Schulz K, Altman D, Moher D, CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMC Med 2010 Mar 24;8:18 [FREE Full text] [doi: 10.1186/1741-7015-8-18] [Medline: 20334633]
12. Pont SJ, Puhl R, Cook SR, Slusser W, SECTION ON OBESITY, OBESITY SOCIETY. Stigma Experienced by Children and Adolescents With Obesity. Pediatrics 2017 Dec;140(6):e20173034. [doi: 10.1542/peds.2017-3034] [Medline: 29158228]
13. Jones B, Madden G, Wengreen H. The FIT Game: preliminary evaluation of a gamification approach to increasing fruit and vegetable consumption in school. Prev Med 2014 Nov;68:76-79. [doi: 10.1016/j.ypmed.2014.04.015] [Medline: 24768916]
14. Cafazzo JA, Casselman M, Hamming N, Katzman DK, Palmert MR. Design of an mHealth app for the self-management of adolescent type 1 diabetes: a pilot study. J Med Internet Res 2012 May 08;14(3):e70 [FREE Full text] [doi: 10.2196/jmir.2058] [Medline: 22564332]
15. Chu JTW, Wan A, Stewart SM, Ng KT, Lam TH, Chan SS. Recruitment and Lessons Learned from a Community-Based Intervention Program: The Learning Families Project in Hong Kong. Front Public Health 2018;6:16 [FREE Full text] [doi: 10.3389/fpubh.2018.00016] [Medline: 29450196]
16. Hamerton H, Mercer C, Riini D, McPherson B, Morrison L. Evaluating Maori community initiatives to promote healthy eating, healthy action. Health Promot Int 2014 Mar;29(1):60-69. [doi: 10.1093/heapro/das048] [Medline: 22952336]
17. Marks A, Wilkes L, Blythe S, Griffiths R. A novice researcher's reflection on recruiting participants for qualitative research. Nurse Res 2017 Sep 19;25(2):34-38. [doi: 10.7748/nr.2017.e1510] [Medline: 29115753]
18. Ford KL, Albritton T, Dunn TA, Crawford K, Neuwirth J, Bull S. Youth Study Recruitment Using Paid Advertising on Instagram, Snapchat, and Facebook: Cross-Sectional Survey Study. JMIR Public Health Surveill 2019 Oct 09;5(4):e14080 [FREE Full text] [doi: 10.2196/14080] [Medline: 31599739]
19. Brauner P, Calero Valdez A, Schroeder U, Ziefle M. Increase Physical Fitness and Create Health Awareness through Exergames and Gamification. In: Holzinger A, Ziefle M, Hitz M, Debevc M, editors. Human Factors in Computing and Informatics. SouthCHI 2013. Lecture Notes in Computer Science, vol 7946. Berlin, Heidelberg: Springer; 2013:349-362.
20. Lentelink S, Spil A, Broens T, Hermens H, Jones V. Healthy weight game!: Lose weight together. 2013 Presented at: 2013 IEEE 2nd International Conference on Serious Games and Applications for Health (SeGAH); May 2-3, 2013; Algarve, Portugal p. 1-8. [doi: 10.1109/segah.2013.6665311]
21. Webb R. New Zealand teachers on strike. Jacobin. 2019. URL: https://jacobinmag.com/2019/06/new-zealand-teachers-strike-budget-wages [accessed 2022-05-17]
22. Baker E. Principles of practice for academic/ practice/community research partnerships. American Journal of Preventive Medicine 1999 Jan;16(1):86-93 [FREE Full text] [doi: 10.1016/s0749-3797(98)00149-4]
23. Kerstetter K. Insider, Outsider, or Somewhere Between: The Impact of Researchers' Identities on the Community-Based Research Process. Journal of Rural Social Sciences 2012;27(2):Article 7 [FREE Full text]
24. Simmons D, Voyle JA. Reaching hard-to-reach, high-risk populations: piloting a health promotion and diabetes disease prevention programme on an urban marae in New Zealand. Health Promot Int 2003 Mar;18(1):41-50. [doi: 10.1093/heapro/18.1.41] [Medline: 12571091]
25. Ni Mhurchu C, Te Morenga L, Tupai-Firestone R, Grey J, Jiang Y, Jull A, et al. A co-designed mHealth programme to support healthy lifestyles in Māori and Pasifika peoples in New Zealand (OL@-OR@): a cluster-randomised controlled trial. The Lancet Digital Health 2019 Oct;1(6):e298-e307 [FREE Full text] [doi: 10.1016/s2589-7500(19)30130-x]
26. School charter overview. New Zealand Ministry of Education. URL: https://www.education.govt.nz/school/schools-planning-and-reporting/charter-overview-and-template/ [accessed 2022-05-17]

Abbreviations

T2D: type 2 diabetes
RCT: randomized controlled trial

Edited by T Leung; submitted 23.08.21; peer-reviewed by M Allman-Farinelli, B Bock; comments to author 10.01.22; revised version received 05.05.22; accepted 09.05.22; published 17.06.22.

Please cite as:
Maddison R, Baghaei N, Calder A, Murphy R, Parag V, Heke I, Dobson R, Marsh S
Feasibility of Using Games to Improve Healthy Lifestyle Knowledge in Youth Aged 9-16 Years at Risk for Type 2 Diabetes: Pilot Randomized Controlled Trial
JMIR Form Res 2022;6(6):e33089
URL: https://formative.jmir.org/2022/6/e33089
doi:10.2196/33089
PMID:35713955

Original Paper

Evaluation of Normalization After Implementation of the Digital Dutch Obstetric Telephone Triage System: Mixed Methods Study With a Questionnaire Survey and Focus Group Discussion

Bernice Engeltjes[1,2], MSc; Ageeth Rosman[2], PhD; Fedde Scheele[1,3], MD, PhD; Christiaan Vis[4,5], MSc; Eveline Wouters[6,7], MD, PhD

[1]Athena Institute for Transdisciplinary Research, Faculty of Science, VU Amsterdam, Amsterdam, Netherlands

[2]Department of Healthcare Studies, Rotterdam University of Applied Sciences, Rotterdam, Netherlands

[3]Department of Healthcare Education, OLVG Teaching Hospital, Amsterdam, Netherlands

[4]Department of Neuro & Development Psychology, Faculty of Behavioral and Movement Sciences, VU Amsterdam, Amsterdam, Netherlands

[5]Amsterdam Public Health Research Institute – Mental Health, Amsterdam, Netherlands

[6]Department of Tranzo, Tilburg School of Social and Behavioral Sciences, Tilburg University, Tilburg, Netherlands

[7]Department of Allied Health Professions, Health Innovations and Technology, Fontys University of Applied Sciences, Eindhoven, Netherlands

Corresponding Author:
Bernice Engeltjes, MSc
Department of Healthcare Studies
Rotterdam University of Applied Sciences
Rochussenstraat 198
Rotterdam, 3015AK
Netherlands
Phone: 31 641804881
Email: engeltjesbernice@gmail.com

Abstract

Background: The Dutch Obstetric Telephone Triage System (DOTTS) was developed to improve the quality of acute obstetric care. To achieve optimal effect, the DOTTS should be adopted in the daily care process by triage staff.

Objective: The primary aim was to evaluate the degree of implementation (ie, normalization) of the DOTTS, and the secondary aim was to evaluate which lessons can be learned from its current implementation in Dutch hospitals.

Methods: An evaluation study with a mixed methods design was performed. All triage staff in 9 Dutch hospitals that implemented the DOTTS before September 1, 2019, were invited to complete the Normalization Measure Development (NoMAD) questionnaire between December 2019 and July 2020. The questionnaire is based on the Normalization Process Theory (NPT). This self-reported questionnaire provides insights into the work people do in order to integrate and embed new practice in routine care. The NPT is based on the following 4 constructs: coherence, cognitive participation, collective action, and reflexive monitoring. Within the questionnaire, each construct is represented by 4-7 questions. Questions are scored on a 5-point normalization process scale. Descriptive statistics were used for analysis of questionnaire scores. Subsequently, the questionnaire findings were discussed during a focus group. Template analysis following the 4 constructs was used for analyzing the results of the focus group.

Results: Overall, 173 of 294 (58.8%) triage staff members completed the NoMAD questionnaire, and 90.2% (156/173) of the participants had used the DOTTS for over 6 months. The digital application was used as much as possible or always by 137 of 173 (79.2%) participants. The overall normalization process score was 3.77 (SD 0.36). The constructs coherence, cognitive participation, collective action, and reflexive monitoring scored 4.01 (SD 0.47), 4.05 (SD 0.45), 3.5 (SD 0.45), and 3.72 (SD 0.47), respectively. Analysis of the focus group discussion showed that the added value of the DOTTS was seen as a quality improvement for the care of pregnant women. Dedication of the complete multidisciplinary implementation team was important for facilitating normalization. Support from the medical staff and proper use by all disciplines involved in the triage were seen as facilitating factors. Participants appreciated training and evaluation, and indicated a need for ongoing training and evaluation in relation to goal achievement.

Conclusions: The DOTTS has been integrated into normal care in daily practice. Evaluation by the NoMAD questionnaire provided a positive overall score. These results are in line with or, in some aspects, better than the results of other evaluation

studies. Key factors in the normalization process of the DOTTS in obstetric triage are the shared added value for stakeholders, the dedication of the complete multidisciplinary implementation team, and implementation plans that are tailor made in the practical context of the hospital.

(JMIR Form Res 2022;6(6):e33709) doi:10.2196/33709

KEYWORDS

obstetric triage; Normalization Process Theory; implementation strategy; hierarchy; medical staff

Introduction

The Dutch Obstetric Telephone Triage System (DOTTS) was developed to provide a uniform and practical basis for estimating the severity of symptoms for unplanned obstetric care requests by telephone. In general, a triage system that prioritizes care according to medical urgency has a favorable effect on the safety and efficiency of emergency care [1,2]. The DOTTS is a reliable [3] and valid [4] evidence-based guideline in which presenting symptoms are used to classify the level of urgency from acute hospital admission using transport by ambulance to self-care with advice at home. It was developed through a multiphase multicenter study in consultation with all relevant stakeholders [5]. The stakeholders can be categorized into nursing, medical, and supporting service personnel. In the first category, we included specialized nurses, general nurses, and doctors' assistants. The second category consisted of obstetricians, obstetricians in training, and midwives. Supporting service personnel consisted of policy makers, managers and management team leaders, and information technology (IT) professionals. All stakeholders were involved in this new activity. The DOTTS has been developed as a digital application and is supported by training of the staff responsible for triage. The DOTTS can be considered as a substantial innovation within the field of obstetric emergency care because it prioritizes care based on the level of urgency in a prestructured manner and not based on the experience of professionals only, it needs the use of digital tools, and it requires changes in the care processes for pregnant women, as well as shifts in roles and responsibilities and improvements in interprofessional collaboration (Multimedia Appendix 1).

Implementation of new innovations in health care should contribute to improve the quality and effectiveness of care [6]. Many innovations are complex and require multiple changes at different levels and by different actors involved in the care processes. When introducing a complex innovation, evaluation of the implementation can optimize this process, and in turn, lessons learned can improve new or further implementation [7,8]. Implementation science has evolved to provide better understanding and explanation of why implementation of innovations succeeds or fails, with the aim to overcome these problems and to improve the methods or the implementation [6]. Numerous theories, models, and taxonomies of implementation have been defined to classify and study implementation [9,10]. To understand the process of implementation, these theoretical approaches can be divided into 3 overarching aims. The first aim is to understand and explain what influences the outcomes of implementation (eg, determinant frameworks, classical theories, and implementation theories). The second aim is to describe and supervise the process of translating research into practice (eg, process models). Finally, the third aim is to evaluate implementation (eg, evaluation frameworks) [11].

In implementation science, attention is paid to the context of implementation [12]. The context can be divided at micro, meso, and macro levels [12-14]. Individual patients and professionals are considered to reflect the micro level. The meso level consists of intraorganizational matters that are characterized by culture and climate, readiness to change, support, and structures within the organization. The macro level is described as the wider environment of exogenous influences, such as policy, guidelines, benchmarking, and the organizational network. Lastly, social relations and support, financial resources, leadership, time availability, evaluation, and physical environment are referred to as being influential at all 3 levels [12-14].

At the organizational level, the Normalization Process Theory (NPT) [15-18] has been developed and added to implementation science. The NPT characterizes implementation as a social process of collective action [15-18]. The NPT offers a framework for process evaluation and for comparative studies of complex interventions. It focuses on factors that promote or inhibit routine embedding of complex interventions in health care practice from a care delivery perspective as opposed to patient- or system-level perspectives. Interactions between the intervention and the way caregivers work in a particular context are seen as core elements of the NPT [15-17]. This theory has been present for some time, has an established scientific basis, and has been evaluated several times for reliability and validity [15-17,19-22]. According to the NPT [15-17], routine use of innovations in practice (ie, the fact that an innovation becomes "normal practice") can be understood in the following 4 constructs: coherence, cognitive participation, collective action, and reflexive monitoring. The construct coherence considers the clarity of the goal and the importance of the intervention for the individual care provider and among care providers jointly. Cognitive participation describes the way that care providers understand and commit to the working method of the intervention, as well as which work processes have changed as a result of the intervention. The construct collective action considers to what extent sufficient support, training, time, and the actual work to carry out the implementation are experienced. The construct reflexive monitoring describes the extent to which the intervention is evaluated and continues to align with expectations, needs, and progressive understanding (ie, reflection). Together, these 4 constructs provide a heuristic tool for understanding and explaining change processes in health care. In the context of understanding the implementation of the

DOTTS within hospitals, we chose to use the NPT as a heuristic tool.

The primary aim of this study was to evaluate the degree of implementation (ie, normalization) of the DOTTS, and the secondary aim was to evaluate which lessons can be learned from its current implementation in Dutch hospitals. This evaluation of the implementation process and the intervention, within the first 9 of 65 (14%) Dutch hospitals, can help to optimize current and new implementations.

Methods

Design

An evaluation study of the implementation of the DOTTS in daily practice with a mixed methods design was performed. As methods, a questionnaire survey and a qualitative focus group discussion were used.

Participating Hospitals and the Context of Implementation

All 9 hospitals that implemented the DOTTS before September 1, 2019, were included in this study. Of the 9 hospitals, 2 were academic hospitals, 5 were teaching hospitals, and 2 were nonteaching hospitals in the Netherlands. Participating hospitals were (1) Erasmus MC Rotterdam, (2) Leiden University Medical Center Leiden, (3) Jeroen Bosch Hospital 's-Hertogenbosch, (4) Antonius Hospital Utrecht, (5) OLVG Amsterdam, (6) Amphia Hospital Breda, (7) Elisabeth Tweesteden Hospital Tilburg, (8) Tjongerschans Hospital Heerenveen, and (9) IJsselland Hospital Capelle aan de IJssel.

In all participating hospitals, the DOTTS was implemented and introduced into routine care. Implementation strategies of the DOTTS were designed for each hospital separately. For this aim, each hospital formed an implementation team with stakeholders. In each hospital, stakeholders involved were at least one nurse and one other care professional (ie, midwife, obstetrician, or obstetrician in training). In most hospitals, implementation teams were much more extensive. The implementation team comprised of a cross-functional team including managers, several nurses, doctors' assistants from the triage department and outpatient clinic, at least two professionals of the medical team, and an IT professional for adding the digital application of the DOTTS into the electronic patient record system. The implementation team jointly developed a tailored implementation plan, which was an actionable specific work plan for the users of each individual hospital. This work plan included the following steps: researching whether there is support for the innovation, performing a baseline measurement, and formulating relevant goals of triage. In addition to the formulated goals, it was important to organize the right facilities, such as a physical workplace for obstetric triage with a computer, telephone, and headset. In addition, an important step during the implementation process was integrating the digital application of the DOTTS into the hospital's electronic patient record system. Importantly, in all hospitals, specific training about the DOTTS was given to the staff responsible for triage. In most hospitals, this training was outsourced to an external organization. In some hospitals, this training was given by in-house experts. This choice was determined by the implementation team. Lastly, providing information to third parties before and after implementation was also an important step. The information about the innovation was given to patients, colleagues, and other cooperation partners (eg, general practitioners and the hospital emergency department).

The implementation team went through the implementation strategies before getting started with the DOTTS. The order, as well as the extent of the steps performed, differed per hospital. Progress was evaluated and adjusted through interim process evaluation. The steps were not taken sequentially. Most implementation teams used process steps, which, in retrospect, show similarities with the process models of Kotter or Grol and Wensing [23-25]. These frameworks are intended to support stepwise planning and management of implementation efforts.

Sample

The participants of this study were users of the DOTTS. Users of the DOTTS were obstetrical nurses, nurses, and doctors' assistants, and they can be seen as triage staff. Users from all hospitals where the DOTTS was implemented were included. Before inviting participants for the questionnaire survey and the focus group discussion, an exploratory meeting was held with the manager of the department where the participants were employed. In this meeting, the manner of invitation of participants was discussed. Hence, a list of participants was formed. According to the preference of the managers and the intended participants, either email addresses were provided or an information letter including a hyperlink to the questionnaire was forwarded to the participants by the manager. At the end of the questionnaire, the participants were asked if they also wanted to participate in a follow-up study (focus group). After consent, these participants were contacted again for participation in the focus group.

Measures and Statistical Analysis

Questionnaire

The validated [26] Dutch version of the Normalization Measure Development (NoMAD) questionnaire based on the conceptual framework of the NPT was used [20,21,27]. Within the NoMAD questionnaire, each construct of the NPT is represented by 4-7 questions. Questions are answered using the normalization process scale (NPS) as follows: 1, not relevant; 2, strongly disagree; 3, disagree; 4, agree; and 5, strongly agree [26].

In addition to the NoMAD questionnaire, 9 questions for characteristics were added to assess representativeness and distributions. These involved age, professional category (ie, obstetrical nurse, nurse, doctor's assistant, or other), type of hospital (ie, academic, teaching, or nonteaching), obstetric experience (ie, years), average hours per week spent on triage activities, number of consultations on average per week, start date of the use of the DOTTS, and frequency of the use of the digital application of the DOTTS. All questions were incorporated into an online questionnaire (Qualtrics [28]). Between December 2019 and July 2020, data were collected during a 3-month period in each hospital.

Analyses of participants' characteristics are presented as numbers or means with percentages or SDs. Descriptive statistics (scale means) were used for the analysis of the questionnaire scores. Analyses of the NPS are presented as numbers with SDs, with minimum and maximum scores. To assess whether the reliability in a different area is sufficient, we also calculated the Cronbach α for the pooled data set. Moreover, the frequency distribution of item responses is presented as the percentage of respondents reporting strongly disagree, disagree, agree, or strongly agree, or respondents who chose to not rate a specific item (not relevant). The questionnaire data analyses were performed in RStudio [29] using psych (scores) [30] and ggplot2 (graphs) [31].

Focus Group

Participants of the focus group were triage staff and users of the DOTTS who had completed the questionnaire. A group discussion was held to triangulate and verify the score of the NoMAD questionnaire and the inhibitory and facilitating factors of DOTTS implementation. Participants were asked to discuss whether they recognized and agreed with subscale scores, and needed to come up with possible explanations about differences per construct. Topics were formed based on organizational context factors [12-14] and were structured following the 4 constructs of the NPT [20-22,26]. Context factors were culture and climate, readiness to change, support, policy, guidelines, benchmarking, organizational network, social relations and support, financial resources, leadership, time availability, feedback, and physical environment [12-14].

In April 2021, a digital focus group (Microsoft Teams) meeting was held, recorded, and transcribed verbatim. Atlas-ti [32] was used during template analysis [33] of the focus group results following the 4 constructs of the NPT [20-22,26]. Member check was performed by all participants. Peer-review template analysis [33] was performed with 3 researchers (BE, EMJW, and ANR).

Ethics Approval

All participants were informed about the study and provided digital informed consent prior to the use of the data for analysis. All data were anonymously processed. Participants were able to withdraw at any time, without any statement of reason. The study was approved by the boards of the Medical Research Ethics Committees United, the Medical Ethics Committee of Leiden University Medical Center, and Erasmus MC of Rotterdam (W.16.053 & P17.075/PG/pg & C1.20191125).

Results

Characteristics of the Participants

In total, 294 triage staff members from the 9 hospitals were asked to complete the questionnaire. The overall response rate, after 3 reminders, for complete responses was 58.8% (173/294).

The participants who filled out the questionnaire had a mean age of 43.3 years (SD 11.6 years) and an average work experience in obstetrics of 17.9 years (SD 11.5 years). Participants in the focus group had a mean age of 46 years (SD 9.2 years) and an average work experience in obstetrics of 18.6 years (SD 9.9 years). An overview of the characteristics of the participants is provided in Table 1. In total, 156 of the 173 (90.2%) participants had used the DOTTS for over 6 months. The digital application of the DOTTS was used "as much as possible" or "always" by 137 of the 173 (79.2%) participants.

Table 1. Characteristics of the participants.

Characteristic	Questionnaire survey (N=173)	Focus group (N=8)
Age (years), mean (SD)[a]	43.3 (11.6)	46.0 (9.2)
Work experience in obstetrics (years), mean (SD)	17.9 (11.5)	18.6 (9.9)
Professional category, n (%)[b]		
Obstetrical nurse	148 (83.1)	7 (87.5)
Nurse	6 (3.5)	0 (0)
Doctor's assistant	11 (6.4)	1 (12.5)
Other	8 (4.6)	0 (0)
Hospital type, n (%)[b]		
Academic hospital	67 (38.7)	2 (25.0)
Teaching hospital	67 (38.7)	4 (50.0)
Nonteaching hospital	39 (22.5)	2 (25.0)
Time performing triage (average) per week, n (%)[b]		
≥16 hours	49 (36.8)	4 (50.0)
9-15 hours	50 (31.6)	1 (12.5)
≤8 hours	74 (31.6)	3 (37.5)
Number of consultations (average) per week, n (%)[b]		
50-100	6 (3.5)	0 (0)
20-49	20 (11.6)	2 (25.0)
10-19	55 (31.8)	3 (37.5)
0-9	90 (52.0)	3 (37.5)
0	2 (1.2)	0 (0)
Duration of use of the DOTTS,[c] n (%)[b]		
≥24 months	29 (16.8)	5 (62.5)
13-24 months	75 (43.3)	3 (37.5)
6-12 months	52 (30.6)	0 (0)
≤6 months	17 (9.8)	0 (0)
Frequency of use of the digital application of the DOTTS, n (%)[b]		
Always	48 (27.7)	3 (37.5)
As much as possible	89 (51.4)	5 (62.5)
Regularly	21 (12.1)	0 (0)
Sometimes	13 (7.5)	0 (0)
Never	2 (1.2)	0 (0)

[a]Missing data (n=1) for age in the questionnaire survey group.

[b]Owing to rounding, the percentages do not add to 100%.

[c]DOTTS: Dutch Obstetric Telephone Triage System.

Results of the Questionnaire Survey

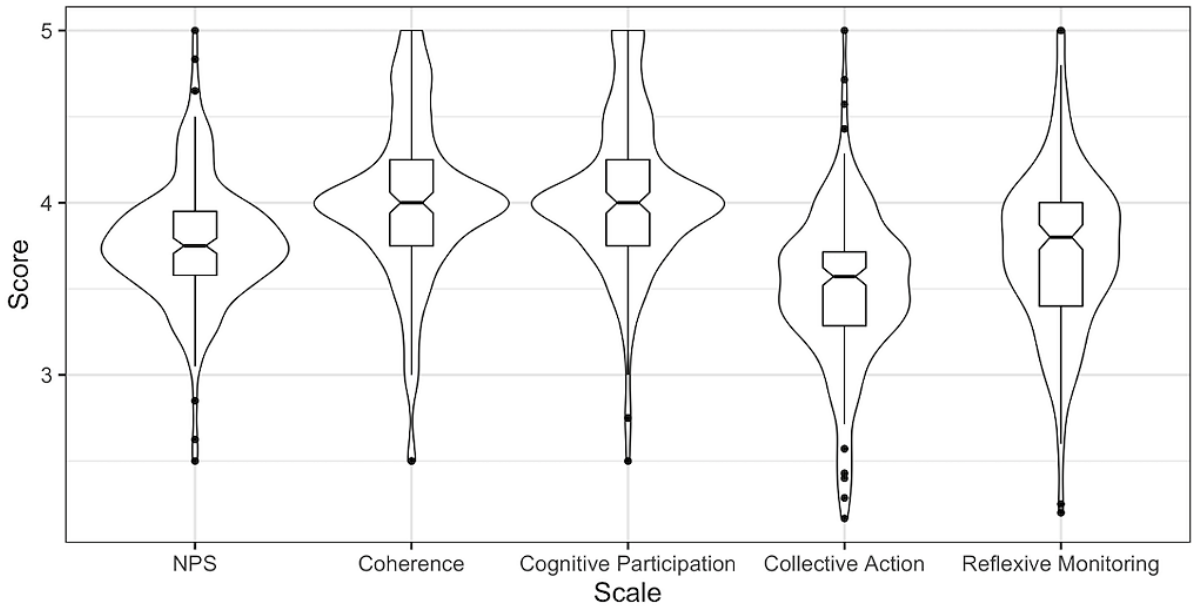
The overall NPS score was 3.77 (SD 0.36). The constructs coherence, cognitive participation, collective action, and reflexive monitoring scored 4.01 (SD 0.47), 4.05 (SD 0.45), 3.5 (SD 0.45), and 3.72 (SD 0.47), respectively (Table 2). On average, all participants agreed (score 4) with the statements associated with the constructs coherence and cognitive participation. For the constructs collective action and reflexive monitoring, the scores were between 4 (agree) and 3 (disagree). These results were also seen when each hospital was analyzed separately (Multimedia Appendix 2). The scores for the constructs collective action and reflexive monitoring showed more variation compared to the scores for the constructs coherence and cognitive participation (Figure 1). All elements

were recognized by most participants. The constructs coherence and cognitive participation had a high percentage of answers with "agree" and "strongly agree" (Multimedia Appendix 3). In the pooled data set, Cronbach α was .85 for the total NPS score and was .71 for coherence, .70 for cognitive participation, .67 for collective action, and .68 for reflexive monitoring (Table 2).

Table 2. Overview of Normalization Measure Development (NoMAD) scale scores (N=173).

NoMAD scale	Mean score (SD)	Score range	Cronbach α
Normalization process	3.77 (0.36)	2.5-5.0	.85
Coherence	4.01 (0.47)	2.5-5.0	.71
Cognitive participation	4.05 (0.45)	2.5-5.0	.70
Collective action	3.50 (0.45)	2.2-5.0	.67
Reflexive monitoring	3.72 (0.47)	2.2-5.0	.68

Figure 1. Box plot of the scale scores of the questionnaires. The results are shown as scale scores (2, strongly disagree; 3, disagree; 4, agree; and 5, strongly agree). NPS: normalization process scale.



Results of the Focus Group Discussion

Eight participants (Table 1) discussed the implementation of the DOTTS in their hospitals and what lessons could be learned. The focus group discussion lasted 90 minutes. The focus group was highly valued by the participants and experienced as a reflection moment. Participants discussed whether they agreed with the NPS and construct scores, and came up with possible explanations.

Coherence

The added value achieved with the implementation of the DOTTS was considered the improvement of the quality of care services for pregnant women. This corresponds to the construct coherence of the NPT. Participants indicated that implementation of obstetric triage provides uniformity in obstetric emergency care, which underpins a quality improvement for the triage ward. This goal was realized with a dedicated and multidisciplinary implementation team, who, in close cooperation with all users, organized and supervised the implementation of the DOTTS. The multidisciplinary team should consist of representatives from the nursing and medical groups, and commitment and support from the manager are also considered important. The implementation team should be able to create sufficient support and ensure joint ownership of the change. Participants indicated that, among other things, good preparation of the team, sufficient description and clarity of roles and responsibilities, and experience in facilitating implementation were important (Multimedia Appendix 4).

Cognitive Participation

To achieve quality improvement, the competencies of triage staff (ie, daily users) should align with the goal of implementation. Dedication, self-efficacy, goal pursuit, and multitasking were mentioned as important competencies to contribute to achieve the added value of the DOTTS. Facilitating factors were clear working agreements for all health care professionals, sufficient capacity of the outpatient clinic organized at the management level, and triage staff who continue to clarify roles and responsibilities of the triage ward with other health care professionals. This corresponds with the construct cognitive participation of the NPT (Multimedia Appendix 4).

Collective Action

The added value of the DOTTS was impeded when there was improper use of the obstetric triage ward. Regular outpatient clinic visits, as opposed to real emergencies, were occasionally allowed to be seen at the obstetric triage ward. The reason for perceived improper use of the triage ward by medical staff, referrers, and staff of the outpatient clinic or labor ward, is associated with several factors, including ambiguity in policy between the triage ward and outpatient clinic or labor ward, a lack of capacity in the outpatient clinic, and a decision by medical staff in a hierarchical manner that a regular appointment is to be made at the triage ward. Where the support of medical staff was lacking, this was, in particular, experienced as an important barrier. However, when medical staff were strategically informed and involved by the representative of the implementation group, this barrier was no longer experienced. These reasons correspond to the construct collective action of the NPT (Multimedia Appendix 4).

Organizing adequate training was perceived as supportive of success. Experiences with training varied among the participants, but on every occasion, training contributed to the understanding and implementation of the DOTTS. The participants stated that ongoing training is a facilitating factor in continuous stimulation of daily use. In addition, nurses who work as triage staff need to be well supported in their new task. In addition to performing obstetric triage, appropriate support services, such as administration and equipment, must be facilitated. Group responsibility for such tasks is necessary to foster ongoing ownership and improvement of the service. Developing a sense of responsibility or co-responsibility for the total organization of care and implementation among triage staff is necessary (Multimedia Appendix 4).

Reflexive Monitoring

The participants discussed the importance of and the amount of regular evaluation for all stakeholders before, during, and after implementation. Within the different hospitals, there were different experiences with the amount and frequency of evaluation. As users of the DOTTS, participants appreciated evaluation and indicated a need for ongoing evaluation in relation to goal achievement. This corresponds to the construct reflexive monitoring of the NPT (Multimedia Appendix 4).

Discussion

Principal Findings

This study aimed to evaluate the use of the DOTTS in daily practice after its implementation in a hospital. Evaluation focused on daily use and on the 4 constructs of the NPS. The DOTTS was used by almost all participants over a period of 6 months or more. The digital application of the DOTTS was used as much as possible or always by most participants. The overall score of the NoMAD questionnaire was 3.77 (SD 0.36). There were some differences per construct, where coherence and cognitive participation scored better and with less variation than collective action and reflexive monitoring. Outcomes of the focus group discussion confirmed the added value of the DOTTS. Use was stimulated by the presence of a dedicated multidisciplinary team and supported by medical staff, as well as proper use of the triage ward, adequate training, and official evaluation.

Comparison With Prior Work

Our results are in line with and, in some aspects, better than the results of other evaluation studies with complex implementations using the NoMAD questionnaire [26,34,35]. While the Dutch questionnaire was previously applied to e-mental health interventions [26], this is the first time it was applied in obstetrics. To assess whether the reliability in a different area of health care is sufficient, we also looked at the Cronbach α. The results of Cronbach α showed good internal consistency for the total NPS score and acceptable findings for coherence and cognitive participation. However, the results were questionable for the constructs collective action and reflexive monitoring. Our findings are comparable to previous results when using the Dutch version of the questionnaire [26].

Triangulation of our results was facilitated via a focus group discussion. What emerged from this discourse was that triage professionals were able to see the added value (coherence) and were committed (cognitive participation), but struggled with collaboration (collective action) to use the DOTTS and did not always reflect on their efforts (reflexive monitoring) in a systematic manner. A plausible explanation for the favorable results of coherence and cognitive participation is the early and intensive involvement of stakeholders in the development and implementation of the DOTTS, which supports implementation. Stakeholders were involved in the development and gave their commitment about the use of the DOTTS in daily practice. The innovation was created together and is therefore well suited to the needs of care providers [5,36-38]. The construct collective action showed the widest variation. One possible explanation for this variation is the tailor-made approach adopted for the implementation plan. The order, as well as the extent of the steps of the implementation plan, differed per hospital. Moreover, the context differed per hospital, which means that every implementation was also different [12].

The lesson learned from this study is that evaluation (ie, reflection) of preplanned, systematic, and strategic implementation of an innovation in health care deserves more attention. In our study, we mainly evaluated whether the use of the DOTTS normalized after implementation (ie, a state of affairs). Each hospital made a tailor-made plan per implementation, which retrospectively showed similarities with the process models of Kotter or Grol and Wensing [6,23-25]. The change management model by Kotter [25] and the model by Grol and Wensing [6] intended to support the planning and managing implementation efforts. In this study, there was no specific model used; therefore, it is difficult to compare the results with these well-known implementation models. However, evaluation of use, similar to the construct reflexive monitoring, is also an important step in these models.

Improving the quality of services for pregnant women was seen as important after the implementation of the DOTTS. Improvement of quality is mentioned in most implementation science research as a condition for success [14]. Improper use of the triage ward, which was considered by participants as a

barrier, is also a well-known phenomenon in the organization of care within the hospital and specifically in triage wards [38-40]. Commonly, in planned hospital care, capacity is limited, causing nonurgent care events to be diverted to the emergency care department. Several factors leading to improper use were also mentioned in the scoping review of Bailey et al [41]. Improper use brings challenges, such as workload stress, which subsequently influence decisions during telephone triage [40,41].

In line with the constructs collective action, cognitive participation, and coherence, implementation by a dedicated multidisciplinary implementation team, which provides guidance during implementation and use of the tool afterwards, was mentioned as important. In our study, this referred to the importance of the involvement of all stakeholders. Preparing an innovation with all stakeholders creates the possibility of optimal support from the start and user friendliness for all stakeholders in daily practice [5,14,42]. Within the multidisciplinary team, special attention should be paid to the participation of the medical group. To change medical staff routines, leadership of the implementation team is an important element [43]. Hierarchy is also a challenging factor here, which is in line with results from other studies indicating that in a hierarchical organization, normalization of an innovation is often more difficult [35]. If the organization of care is arranged by the nurse, it is necessary that it is supported by medical staff [40].

We found that training, which is part of the construct collective action, is an important element of change. This was also seen by existing triage systems [38,44,45]. If users themselves have a need for training because they want to be competent, this contributes to success [46]. After completing the training, it is important to provide continuous evaluation, which contributes to the construct reflexive monitoring, so that the implementation is further optimized and users receive confirmation that they are doing well [38].

Strengths, Limitations, and Recommendations

The NPT with its validated NoMAD questionnaire was used as an evaluation framework in this study. The NPT is an implementation theory and has been widely used as an evaluation framework [11]. There are also other tools that reflect the success of implementation, such as the Consolidated Framework for Implementation Research (CFIR); Nonadoption Abandonment, Scale-up, Spread, and Sustainability (NASSS) framework; and Reach, Effectiveness, Adoption, Implementation, Maintenance (RE-AIM) framework [11]. The NPT was used in this study because the context of the implementation of an innovation in obstetric care corresponded to the NPT [15].

The use of 2 complimentary research methods is valuable in the interpretation of data. A large group of users were given the opportunity to evaluate the use of the DOTTS, providing a general overview of implementation in several hospitals. The focus group gave the opportunity to triangulate the outcomes, thereby gaining more insight into the meaning of the answers and clarifying the context.

The results are from all hospitals that implemented the DOTTS before September 2019. There was an overall response rate of 58.8% (173/294) from these 9 hospitals. A 50% response rate, which was obtained from all hospitals, can be considered representative [47] (Multimedia Appendix 2). With an average age of 43.3 years and experience of 17.9 years, the sample composition was representative compared to other studies within this profession [48-50]. The participants of the focus group showed good representation of the total research group (Table 1).

To improve the questionnaire, it is recommended to look to the question collective action-2 of the construct collective action because it is the only negatively asked question. It is unclear if every participant interpreted the question correctly.

In this study, we chose to evaluate the degree of normalization after implementation among daily users in 9 hospitals where the DOTTS is offered as usual care. This focus resulted in a lack of information about the theoretical approaches used for each implementation strategy. The tailored implementation strategy created space for context per hospital and the team of stakeholders. However, it limited the ability to evaluate effectiveness per implementation strategy. In addition, due to the current aim of this study, results on the expected quality improvement were lacking. Furthermore, this study only looked at the perspectives of the daily users of the DOTTS. The lack of perspectives of medical staff, outpatient clinic staff, management, and other related professionals is a potential limitation. In view of the importance of tailored implementation strategies, which was highlighted by our research, we recommend that a future study should include representation from the medical group to ensure an inclusive perspective.

Moreover, we did not evaluate the patient perspective. Not every patient will fit into the evidence-based system of the DOTTS, for instance, patients who do not understand self-care with advice. Customization per patient might need to be further developed. The current evaluation was not about these items and requires further research. Further insight into the experiences of patients who have received telephone and physical triage care based on the DOTTS is therefore recommended.

Conclusions

Normalization of the DOTTS was seen after tailored implementation in 9 hospitals. Key factors in the normalization process of the DOTTS in obstetric triage were (1) the shared added value for stakeholders; (2) the dedication of the complete multidisciplinary implementation team with specific support from medical staff, as well as proper use of the triage ward (as designed) by all disciplines; and (3) implementation plans that are tailor made in the practical context of the hospital. Improvement can be achieved by structuring this process and incorporating implementation strategies, such as systematic training and evaluation, with users.

Acknowledgments

The authors wish to thank all participating obstetric professionals. The authors also wish to thank Dr DJ Cronie for feedback and editing assistance.

Authors' Contributions

All authors made significant contributions to the work reported (conception, study design, execution, acquisition of data, analysis, and interpretation) and took part in drafting, revising, or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and have agreed to be accountable for all aspects of the work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Video of the Dutch Obstetric Telephone Triage System.
[MP4 File (MP4 Video), 18561 KB - formative_v6i6e33709_app1.mp4]

Multimedia Appendix 2
Subanalysis questionnaire findings per hospital.
[DOCX File , 15 KB - formative_v6i6e33709_app2.docx]

Multimedia Appendix 3
Frequency distribution of item responses. The constructs are coherence (CO), cognitive participation (CP), collective action (CA), and reflexive monitoring (RM). The upper part of the figure shows the percentage of respondents reporting strongly disagree, disagree, agree, or strongly agree. The gray bar coupled to the y-axis indicates the percentage of participants rating an item as "neutral." The lower part of the figure shows the percentage of respondents who chose to not rate a specific item (not relevant).
[PNG File , 266 KB - formative_v6i6e33709_app3.png]

Multimedia Appendix 4
Quotes from the focus group discussion. To illustrate the focus group discussion, several quotes per construct from participants are shown. These quotes have been selected by the authors after analysis.
[DOCX File , 14 KB - formative_v6i6e33709_app4.docx]

References

1. Zachariasse JM, van der Hagen V, Seiger N, Mackway-Jones K, van Veen M, Moll HA. Performance of triage systems in emergency care: a systematic review and meta-analysis. BMJ Open 2019 May 28;9(5):e026471 [FREE Full text] [doi: 10.1136/bmjopen-2018-026471] [Medline: 31142524]
2. Lake R, Georgiou A, Li J, Li L, Byrne M, Robinson M, et al. The quality, safety and governance of telephone triage and advice services - an overview of evidence from systematic reviews. BMC Health Serv Res 2017 Aug 30;17(1):614 [FREE Full text] [doi: 10.1186/s12913-017-2564-x] [Medline: 28854916]
3. Engeltjes B, Rosman A, Bertens LC, Wouters E, Cronie D, Scheele F. Reliability of Dutch obstetric telephone triage. RMHP 2021 Aug;Volume 14:3247-3254. [doi: 10.2147/rmhp.s319564] [Medline: 34393531]
4. Engeltjes B, van Dijk C, Rosman A, Rijke R, Scheele F, Wouters E. Validation of Dutch obstetric telephone triage system: A prospective validation study. RMHP 2021 May;Volume 14:1907-1915. [doi: 10.2147/rmhp.s306390] [Medline: 34007228]
5. Engeltjes B, Wouters E, Rijke R, Scheele F. Obstetric telephone triage. RMHP 2020 Nov;Volume 13:2497-2506. [doi: 10.2147/rmhp.s277464] [Medline: 33177905]
6. Wensing M, Grol R. Knowledge translation in health: how implementation science could contribute more. BMC Med 2019 May 07;17(1):88 [FREE Full text] [doi: 10.1186/s12916-019-1322-9] [Medline: 31064388]
7. Perry CK, Damschroder LJ, Hemler JR, Woodson TT, Ono SS, Cohen DJ. Specifying and comparing implementation strategies across seven large implementation interventions: a practical application of theory. Implement Sci 2019 Mar 21;14(1):32 [FREE Full text] [doi: 10.1186/s13012-019-0876-4] [Medline: 30898133]
8. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 07;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
9. Moullin JC, Dickson KS, Stadnick NA, Albers B, Nilsen P, Broder-Fingert S, et al. Ten recommendations for using implementation frameworks in research and practice. Implement Sci Commun 2020 Apr 30;1(1):42 [FREE Full text] [doi: 10.1186/s43058-020-00023-7] [Medline: 32885199]

10. Armstrong R, Sales A. Welcome to implementation science communications. Implement Sci Commun 2020;1:1 [FREE Full text] [doi: 10.1186/s43058-020-00018-4] [Medline: 32885176]
11. Nilsen P. Making sense of implementation theories, models and frameworks. Implement Sci 2015 Apr 21;10:53 [FREE Full text] [doi: 10.1186/s13012-015-0242-0] [Medline: 25895742]
12. Nilsen P, Bernhardsson S. Context matters in implementation science: a scoping review of determinant frameworks that describe contextual determinants for implementation outcomes. BMC Health Serv Res 2019 Mar 25;19(1):189 [FREE Full text] [doi: 10.1186/s12913-019-4015-3] [Medline: 30909897]
13. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. Milbank Q 2004;82(4):581-629 [FREE Full text] [doi: 10.1111/j.0887-378X.2004.00325.x] [Medline: 15595944]
14. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: A new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
15. May CR, Johnson M, Finch T. Implementation, context and complexity. Implement Sci 2016 Oct 19;11(1):141 [FREE Full text] [doi: 10.1186/s13012-016-0506-3] [Medline: 27756414]
16. May C. Towards a general theory of implementation. Implement Sci 2013 Feb 13;8:18 [FREE Full text] [doi: 10.1186/1748-5908-8-18] [Medline: 23406398]
17. May CR, Mair F, Finch T, MacFarlane A, Dowrick C, Treweek S, et al. Development of a theory of implementation and integration: Normalization Process Theory. Implement Sci 2009 May 21;4:29 [FREE Full text] [doi: 10.1186/1748-5908-4-29] [Medline: 19460163]
18. Heinsch M, Wyllie J, Carlson J, Wells H, Tickner C, Kay-Lambkin F. Theories informing eHealth implementation: Systematic review and typology classification. J Med Internet Res 2021 May 31;23(5):e18500 [FREE Full text] [doi: 10.2196/18500] [Medline: 34057427]
19. Murray E, Treweek S, Pope C, MacFarlane A, Ballini L, Dowrick C, et al. Normalisation process theory: a framework for developing, evaluating and implementing complex interventions. BMC Med 2010 Oct 20;8:63 [FREE Full text] [doi: 10.1186/1741-7015-8-63] [Medline: 20961442]
20. Finch TL, Girling M, May CR, Mair FS, Murray E, Treweek S, et al. Improving the normalization of complex interventions: part 2 - validation of the NoMAD instrument for assessing implementation work based on normalization process theory (NPT). BMC Med Res Methodol 2018 Nov 15;18(1):135 [FREE Full text] [doi: 10.1186/s12874-018-0591-x] [Medline: 30442094]
21. Rapley T, Girling M, Mair FS, Murray E, Treweek S, McColl E, et al. Improving the normalization of complex interventions: part 1 - development of the NoMAD instrument for assessing implementation work based on normalization process theory (NPT). BMC Med Res Methodol 2018 Nov 15;18(1):133 [FREE Full text] [doi: 10.1186/s12874-018-0590-y] [Medline: 30442093]
22. Elf M, Nordmark S, Lyhagen J, Lindberg I, Finch T, Åberg AC. The Swedish version of the Normalization Process Theory Measure S-NoMAD: translation, adaptation, and pilot testing. Implement Sci 2018 Dec 04;13(1):146 [FREE Full text] [doi: 10.1186/s13012-018-0835-5] [Medline: 30509289]
23. Wensing M, Bosch M, Grol R. Developing and selecting interventions for translating knowledge to action. CMAJ 2010 Feb 09;182(2):E85-E88 [FREE Full text] [doi: 10.1503/cmaj.081233] [Medline: 20026633]
24. Grol R, Wensing M, Eccles M, Davis D. Improving Patient Care: The Implementation of Change in Health Care, 2nd Edition. Hoboken, New Jersey: Wiley-Blackwell; 2013:1-392.
25. Campbell RJ. Change management in health care. Health Care Manag 2020;39(2):50-65. [doi: 10.1097/hcm.0000000000000290]
26. Vis C, Ruwaard J, Finch T, Rapley T, de Beurs D, van Stel H, et al. Toward an objective assessment of implementation processes for innovations in health care: Psychometric evaluation of the Normalization Measure Development (NoMAD) questionnaire among mental health care professionals. J Med Internet Res 2019 Feb 20;21(2):e12376 [FREE Full text] [doi: 10.2196/12376] [Medline: 30785402]
27. Normalization Measure Development Questionnaire (NoMAD). Implement All. 2018. URL: https://www.implementall.eu/NoMAD_English_2.pdf [accessed 2022-06-02]
28. Qualtrics. URL: https://www.qualtrics.com [accessed 2022-06-02]
29. RStudio. URL: http://www.rstudio.com/ [accessed 2022-06-02]
30. Procedures for Psychological, Psychometric, and Personality Research. R Project. URL: https://cran.r-project.org/web/packages/psych/psych.pdf [accessed 2022-06-02]
31. Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York, New York: Springer-Verlag; 2016:1-260.
32. ATLAS.ti Scientific Software Development. URL: https://atlasti.com/ [accessed 2022-06-02]
33. Brooks J, McCluskey S, Turley E, King N. The utility of template analysis in qualitative psychology research. Qual Res Psychol 2015 Apr 03;12(2):202-222 [FREE Full text] [doi: 10.1080/14780887.2014.955224] [Medline: 27499705]

34. Cook C, Finch T, Sharma S, Girling M, Rapley T, Vernazza CR. Developing oral health risk assessment as routine practice during early stages of clinical careers: A cross-sectional study of dental students using the NoMAD questionnaire. Eur J Dent Educ 2020 Feb 10;24(1):169-176. [doi: 10.1111/eje.12481] [Medline: 31765500]
35. May CR, Cummings A, Girling M, Bracher M, Mair FS, May CM, et al. Using Normalization Process Theory in feasibility studies and process evaluations of complex healthcare interventions: a systematic review. Implement Sci 2018 Jun 07;13(1):80 [FREE Full text] [doi: 10.1186/s13012-018-0758-1] [Medline: 29879986]
36. Carr D, Howells A, Chang M, Hirji N, English A. An integrated approach to stakeholder engagement. Healthc Q 2009 May 15;12 Spec No Ontario(sp):62-70 [FREE Full text] [doi: 10.12927/hcq.2009.20754] [Medline: 19458512]
37. Coiera E. Four rules for the reinvention of health care. BMJ 2004 May 15;328(7449):1197-1199 [FREE Full text] [doi: 10.1136/bmj.328.7449.1197] [Medline: 15142933]
38. Moudi A, Iravani M, Najafian M, Zareiyan A, Forouzan A, Mirghafourvand M. Factors influencing the implementation of obstetric triage: a qualitative study. Midwifery 2021 Jan;92:102878. [doi: 10.1016/j.midw.2020.102878] [Medline: 33161173]
39. Reay G, Smith-MacDonald L, Then KL, Hall M, Rankin JA. Triage emergency nurse decision-making: Incidental findings from a focus group study. Int Emerg Nurs 2020 Jan;48:100791. [doi: 10.1016/j.ienj.2019.100791] [Medline: 31494074]
40. Bailey CM, Newton JM, Hall HG. Telephone triage in midwifery practice: A cross-sectional survey. Int J Nurs Stud 2019 Mar;91:110-118. [doi: 10.1016/j.ijnurstu.2018.11.009] [Medline: 30682631]
41. Bailey CM, Newton JM, Hall HG. Telephone triage and midwifery: A scoping review. Women Birth 2018 Oct;31(5):414-421. [doi: 10.1016/j.wombi.2017.12.002] [Medline: 29241698]
42. Nadav J, Kaihlanen A, Kujala S, Laukka E, Hilama P, Koivisto J, et al. How to implement digital services in a way that they integrate into routine work: Qualitative interview study among health and social care professionals. J Med Internet Res 2021 Dec 01;23(12):e31668 [FREE Full text] [doi: 10.2196/31668] [Medline: 34855610]
43. Greenhalgh T, Rosen R, Shaw SE, Byng R, Faulkner S, Finlay T, et al. Planning and evaluating remote consultation services: A new conceptual framework incorporating complexity and practical ethics. Front Digit Health 2021 Aug 13;3:726095 [FREE Full text] [doi: 10.3389/fdgth.2021.726095] [Medline: 34713199]
44. Quaile H. Implementing an obstetrics-specific triage acuity tool to increase nurses' knowledge and improve timeliness of care. Nurs Womens Health 2018 Aug;22(4):293-301. [doi: 10.1016/j.nwh.2018.05.002] [Medline: 30077235]
45. Kenyon S, Hewison A, Dann S, Easterbrook J, Hamilton-Giachritsis C, Beckmann A, et al. The design and implementation of an obstetric triage system for unscheduled pregnancy related attendances: a mixed methods evaluation. BMC Pregnancy Childbirth 2017 Sep 18;17(1):309 [FREE Full text] [doi: 10.1186/s12884-017-1503-5] [Medline: 28923021]
46. Ntoumanis N, Ng JYY, Prestwich A, Quested E, Hancox JE, Thøgersen-Ntoumani C, et al. A meta-analysis of self-determination theory-informed intervention studies in the health domain: effects on motivation, health behavior, physical, and psychological health. Health Psychol Rev 2021 Jun;15(2):214-244. [doi: 10.1080/17437199.2020.1718529] [Medline: 31983293]
47. Cooper AL, Brown J. Maximising nurses' and midwives' response rates to surveys. Nurse Res 2017 Dec 18;25(3):31-35. [doi: 10.7748/nr.2017.e1494] [Medline: 29251446]
48. Perdok H, Cronie D, van der Speld C, van Dillen J, de Jonge A, Rijnders M, et al. Experienced job autonomy among maternity care professionals in The Netherlands. Midwifery 2017 Nov;54:67-72. [doi: 10.1016/j.midw.2017.07.015] [Medline: 28850826]
49. Cronie D, Rijnders M, Jans S, Verhoeven CJ, de Vries R. How good is collaboration between maternity service providers in the Netherlands? JMDH 2018 Dec;Volume 12:21-30. [doi: 10.2147/jmdh.s179811]
50. Nowrouzi B, Lightfoot N, Carter L, Larivière M, Rukholm E, Schinke R, et al. The relationship between quality of work life and location of cross-training among obstetric nurses in urban northeastern Ontario, Canada: A population-based cross sectional study. Int J Occup Med Environ Health 2015;28(3):571-586 [FREE Full text] [doi: 10.13075/ijomeh.1896.00443] [Medline: 26190732]

Abbreviations

DOTTS: Dutch Obstetric Telephone Triage System
IT: information technology
NoMAD: Normalization Measure Development
NPS: normalization process scale
NPT: Normalization Process Theory

Edited by A Mavragani; submitted 20.09.21; peer-reviewed by J Roca, A Joseph; comments to author 29.12.21; revised version received 17.02.22; accepted 26.05.22; published 17.06.22.

Please cite as:
Engeltjes B, Rosman A, Scheele F, Vis C, Wouters E
Evaluation of Normalization After Implementation of the Digital Dutch Obstetric Telephone Triage System: Mixed Methods Study With a Questionnaire Survey and Focus Group Discussion
JMIR Form Res 2022;6(6):e33709
URL: https://formative.jmir.org/2022/6/e33709
doi:10.2196/33709
PMID:35616176

Original Paper

Implementation Outcomes Assessment of a Digital Clinical Support Tool for Intrapartum Care in Rural Kenya: Observational Analysis

Nhi Dinh[1], MSPH; Smisha Agarwal[1], MPH, MBA, PhD; Lisa Avery[2], MIH, MD; Priya Ponnappan[3], MBA; Judith Chelangat[4], BBMIT; Paul Amendola[3], MPH; Alain Labrique[1], MHS, PhD; Linda Bartlett[1], MD, MHSc

[1]Department of International Health, Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, United States

[2]Centre for Global Public Health, University of Manitoba, Winnipeg, MB, Canada

[3]Vecna Cares, Burlington, MA, United States

[4]Vecna Cares, Kilgoris, Kenya

Corresponding Author:
Linda Bartlett, MD, MHSc
Department of International Health
Johns Hopkins Bloomberg School of Public Health
Johns Hopkins University
615 N Wolfe St
Baltimore, MD, 21205
United States
Phone: 1 410 510 1055
Email: linda.bartlett@jhu.edu

Abstract

Background: iDeliver, a digital clinical support system for maternal and neonatal care, was developed to support quality of care improvements in Kenya.

Objective: Taking an implementation research approach, we evaluated the adoption and fidelity of iDeliver over time and assessed the feasibility of its use to provide routine Ministry of Health (MOH) reports.

Methods: We analyzed routinely collected data from iDeliver, which was implemented at the Transmara West Sub-County Hospital from December 2018 to September 2020. To evaluate its adoption, we assessed the proportion of actual facility deliveries that was recorded in iDeliver over time. We evaluated the fidelity of iDeliver use by studying the completeness of data entry by care providers during each stage of the labor and delivery workflow and whether the use reflected iDeliver's envisioned function. We also examined the data completeness of the maternal and neonatal indicators prioritized by the Kenya MOH.

Results: A total of 1164 deliveries were registered in iDeliver, capturing 45.31% (1164/2569) of the facility's deliveries over 22 months. This uptake of registration improved significantly over time by 6.7% (SE 2.1) on average in each quarter-year (P=.005), from 9.6% (15/157) in the fourth quarter of 2018 to 64% (235/367) in the third quarter of 2020. Across iDeliver's workflow, the overall completion rate of all variables improved significantly by 2.9% (SE 0.4) on average in each quarter-year (P<.001), from 22.25% (257/1155) in the fourth quarter of 2018 to 49.21% (8905/18,095) in the third quarter of 2020. Data completion was highest for the discharge-labor summary stage (16,796/23,280, 72.15%) and lowest for the labor signs stage (848/5820, 14.57%). The completion rate of the key MOH indicators also improved significantly by 4.6% (SE 0.5) on average in each quarter-year (P<.001), from 27.1% (69/255) in the fourth quarter of 2018 to 83.75% (3346/3995) in the third quarter of 2020.

Conclusions: iDeliver's adoption and data completeness improved significantly over time. The assessment of iDeliver' use fidelity suggested that some features were more easily used because providers had time to enter data; however, there was low use during active childbirth, which is when providers are necessarily engaged with the woman and newborn. These insights on the adoption and fidelity of iDeliver use prompted the team to adapt the application to reflect the users' culture of use and further improve the implementation of iDeliver.

(JMIR Form Res 2022;6(6):e34741) doi:10.2196/34741

KEYWORDS

newborn; neonatal health; maternal health; intrapartum care; labor and delivery; Kenya; digital clinical decision support; health information systems; digital health; implementation research

Introduction

Background

Kenya has made major strides in preventing maternal deaths, reducing the maternal mortality ratio by half (52%) from 708 maternal deaths per 100,000 livebirths in 2000 to 342 maternal deaths per 100,000 livebirths in 2017 [1]. To promote the use of maternity services and reduce pregnancy-related mortality, a maternal health care policy was implemented in 2013 to abolish fees associated with childbirth in all public health facilities, which resulted in an increase in facility-based deliveries by 29.5%, from 234,601 deliveries before policy implementation to 303,705 deliveries after policy implementation [2]. However, this improved coverage of hospital-based births alone did not yield as much gain in maternal and newborn survival as expected, spurring additional efforts to improve quality of care and achieve Sustainable Development Goal 3.1—to reduce maternal mortality to ≤140 per 100,000 live births by 2030 [2-4].

The first Kenya Ministry of Health (MOH) report on Confidential Enquiry into Maternal Deaths noted that constraints to quality care that were contributing factors to most maternal deaths included delays in initiation of treatment, inadequate clinical skills, insufficient quality monitoring, and poor record keeping and documentation [5,6]. These issues are heavily affected by the underlying factors of understaffing with its associated burnout and fatigue and lack of adequate resources including space, privacy, training, and commodities [7]. Addressing these challenges could improve maternal and neonatal outcomes in Kenya and countries with similar challenges and resource constraints. Bhutta et al [8] reported that among all women giving birth in health facilities worldwide, if 90% of them actually received the recommended interventions during labor and delivery, an estimated 84% (113,000) of maternal deaths, 76% (531,000) of stillbirths, and 77% (1.325 million) of neonatal deaths can be prevented globally. This is supported by other studies reporting that interventions delivered during labor and childbirth provided the maximum benefits to avert stillbirths and neonatal and maternal deaths [9-11]. Similar projections based on Kenya's data from 1990 to 2015 showed that active management of the third stage of labor and treatment of eclampsia accounted for 86% of maternal lives saved, and optimal care during childbirth and the postnatal period accounted for 70% of neonatal deaths averted [12].

Digital health interventions have the potential to improve adherence to recommended protocols by care providers during labor and delivery and to positively impact maternal and neonatal health outcomes [13]. A study on skilled birth attendants in Kenya showed that users of electronic partographs, in comparison with users of paper partographs, were more likely to be compliant with a set of standard practices during labor and delivery, including measuring pulse, temperature, amniotic fluid status, molding, blood pressure, and urine [14]. The use of electronic partographs was also associated with 56% reduction in suboptimal newborn outcomes [14]. According to the World Health Organization's (WHO) recommendations on digital interventions for health system strengthening, the digital monitoring of clients' health status and health service use has the potential to improve continuity and timely provision of care and adherence to clinical guidelines [15].

Objective

Within this context, formative research was undertaken to assess the need and design of a potential digital solution, resulting in the development of iDeliver—a clinical decision support intervention developed by Vecna Cares, Johns Hopkins Bloomberg School of Public Health, and Scope (formerly known as M4ID) and supported by Merck for Mothers. Guided by human-centered design priorities elicited from stakeholders, the intended use of iDeliver was to help health care workers navigate the assessment and triage of pregnant clients on arrival at a facility, guide clinical decision-making during labor and delivery and immediately after childbirth, and streamline data reporting processes. If used correctly and efficiently, iDeliver has the potential to strengthen the quality of pregnancy, childbirth, and newborn care and facilitate the improvement of maternal and newborn health outcomes. This study takes an implementation research approach to evaluate the adoption and fidelity of iDeliver.

Using the iDeliver digital platform implemented at the Transmara West Sub-County Hospital in Kenya, we leveraged routinely collected data in a novel way to assess, quantify, and drive improvements in the *actual use* of this digital intervention by examining the completeness of the data collected. Although noted as a highly useful step in the Monitoring and Evaluating Digital Interventions guide [16], mining routinely collected data is a neglected area of digital implementation—a critical *middle ground* between the frequently assessed *acceptability* and *impact* of digital systems [16]. Acceptability is a necessity, but needs to be translated to use, which is the mediator of impact.

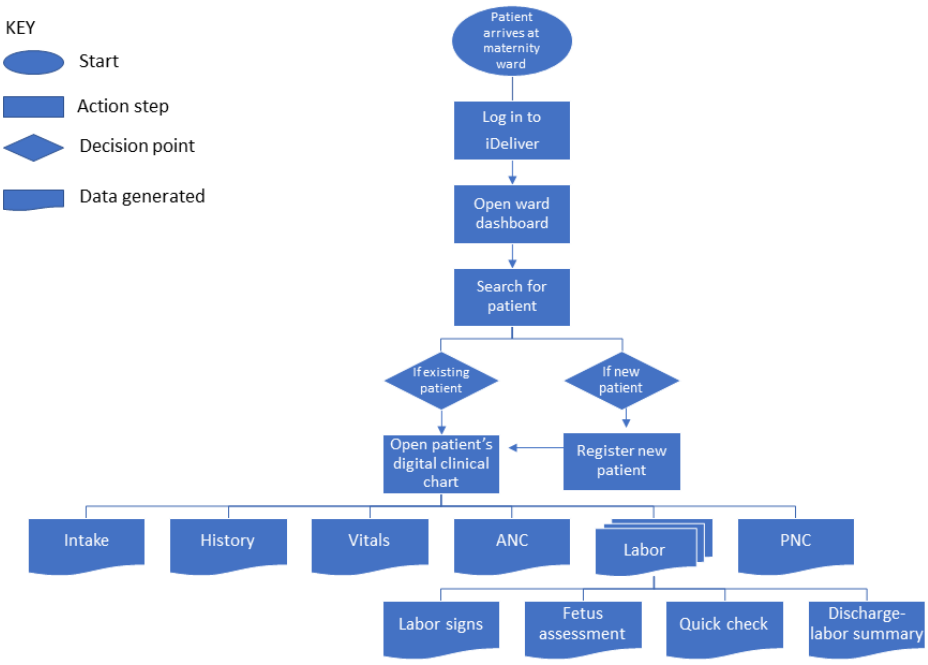
For the purpose of this study, adoption was defined as the uptake of iDeliver [17], and fidelity was defined as the degree to which iDeliver was used as it was designed originally [17], following the definition of implementation outcome variables described by Peters et al [17]. We also examined the data completeness of priority indicators to monitor maternal and newborn health outcomes and quality of care identified by the Kenya MOH's Reproductive and Maternal Health Services [18-21] to assess the feasibility of using iDeliver to provide routine reports to the MOH. This study provided insights on variability in iDeliver use over time and identified areas for improvements and successful deployment strategies. Through this paper, we present a proof of concept of using a decision support tool to support documentation and clinical decision-making around prenatal, intrapartum, and postnatal care. In the real-world setting with limited resources, this process of evaluating implementation outcomes is a crucial intermediate step for a novel digital health intervention such as iDeliver to achieve its ultimate goal of improving maternal and newborn health outcomes. The methods used and corresponding findings could be helpful to other digital health solution implementors during similar early stages of technology deployment.

Methods

iDeliver—Description of the Technology

iDeliver is a software application that allows health care providers to document relevant patient information and clinical progression throughout the continuum of maternal care in real time. Figure 1 summarizes the overall workflow. The version of iDeliver assessed in this study focused on intrapartum care; recent updates to the application also include antenatal and postnatal care components. The health care provider registers a new patient when she arrives at the labor and delivery ward and enters the key patient demographic and clinical information, which generates an acuity score for triage priority. All active registered patients can be seen on a dashboard from which users can access a patient's digital clinical chart, navigate to any section—intake, history, vital signs, labor signs, fetal assessment, and discharge-labor summary—and enter the patient's information at successive appointments to maintain a longitudinal health record. Digital clinical decision support algorithms and patient management guidelines for iDeliver are based on WHO's Managing Complications in Pregnancy and Childbirth [22], Better Outcomes in Labour Difficulty Initiative [23], and Recommendations for Intrapartum Care for a Positive Childbirth Experience [24]. In addition, iDeliver includes clinical training resources, electronic medical record function, and report generation. Further information on the design, development, and implementation of iDeliver has been presented elsewhere [25].

Figure 1. User's workflow through iDeliver when a patient arrives at the maternity ward for labor and delivery. ANC: antenatal care; PNC: postnatal care.



Site Selection

iDeliver was developed in collaboration with nurses, midwives, physicians, and public health administrators in Transmara West and Transmara East Sub-Counties of Narok County, Kenya. It was first implemented in 2017 at the Transmara West Sub-County Hospital in Kilgoris, which is a level-4 tertiary facility offering comprehensive emergency obstetric and newborn care services, with an average of 1150 births annually. It has since been scaled up to 13 other sites in Kenya and Tanzania. This study focused on iDeliver implementation at the Transmara West Sub-County Hospital during the 22-month period after transition to OpenMRS (OpenMRS Inc) platform (December 2018 to September 2020).

Technology Introduction and Training

Since deployment, the application underwent significant updates. Transition from a proprietary to an open-source back end—OpenMRS—was done in November 2018. iDeliver interfaces were built as modular, encapsulated setup code built upon the OpenMRS application platform using ReactJS (Meta), a modern front-end language.

As of September 2020, 5 physicians and 14 nursing officers at the Transmara West Sub-County Hospital were trained to use iDeliver, with 40% (2/5) of the physicians and 57% (8/14) of the nursing officers as current active users. User training was conducted on site every 6 months to account for any upgrades in the application and for staff rotation. Training of new staff occurred on an as-needed basis.

Analytic Approach

Data were extracted and deidentified using MySQL (version 5.6.49; Oracle Corporation). Then, MySQL Workbench (version 8.0; Oracle Corporation) was used to export the data into Excel format. All statistical analyses were performed using R (version 4.0.2; R Foundation for Statistical Computing).

We conducted descriptive analysis to summarize the characteristics of all mothers and newborn infants with information registered in iDeliver within the study period. Then,

we assessed iDeliver's adoption by exploring the following: (1) what proportion of services provided at the health facility are captured by iDeliver and (2) does the uptake of iDeliver use improve over time? To answer the first question and measure iDeliver's uptake, we divided the number of deliveries registered in iDeliver by the number of deliveries recorded on paper at the Transmara West Sub-County Hospital from December 2018 to September 2020. To answer the second question, we assessed the trends in the uptake of iDeliver, by quarter-year, using simple linear regression. A *P* value of <.05 was considered as statistically significant.

We assessed the fidelity of iDeliver use to its original purpose as a decision-making and data management tool by examining which feature or features of iDeliver are used most by users, as assessed by data completion. We used the proportion of data available across the labor and delivery workflow to identify both areas of high use and missed opportunities for use. In particular, we assessed data completion for each stage of the iDeliver's labor and delivery workflow: (1) intake, (2) history, (3) vital signs, (4) labor signs, (5) fetus assessment, (6) quick check, and (7) discharge-labor summary to identify the aspects of the intrapartum process that were plausible for care providers to use and if the use reflected iDeliver's envisioned function for intrapartum clinical guidance. We also assessed the data completion for each stage over time, by quarter-year, using simple linear regression. A *P* value of <.05 was considered as statistically significant.

To assess the feasibility of using iDeliver to provide routine reports based on the priority indicators to monitor maternal and newborn health outcomes and quality of care identified by the Kenya MOH's Reproductive and Maternal Health Services [18-21], we also examined the data completeness of those indicators from the MOH's maternal and perinatal notification and review forms that overlapped with the data in iDeliver. These indicators were referral information (referral from community unit or health facility or referral out to community unit); mother's HIV status; parity; fetal presentation; mode of delivery; date and time of delivery; sex of baby; condition of baby at birth; appearance, pulse, grimace, activity, and respiration score (at 1, 5, and 10 minutes); baby given tetracycline; condition of mother; and condition of baby at discharge. We also assessed the data completeness of these indicators over time, by quarter-year, using simple linear regression. A *P* value of <.05 was considered as statistically significant.

Ethics Approval

The study was approved by the institutional review board of Johns Hopkins Bloomberg School of Public Health (protocol code I18203; December 8, 2021).

Results

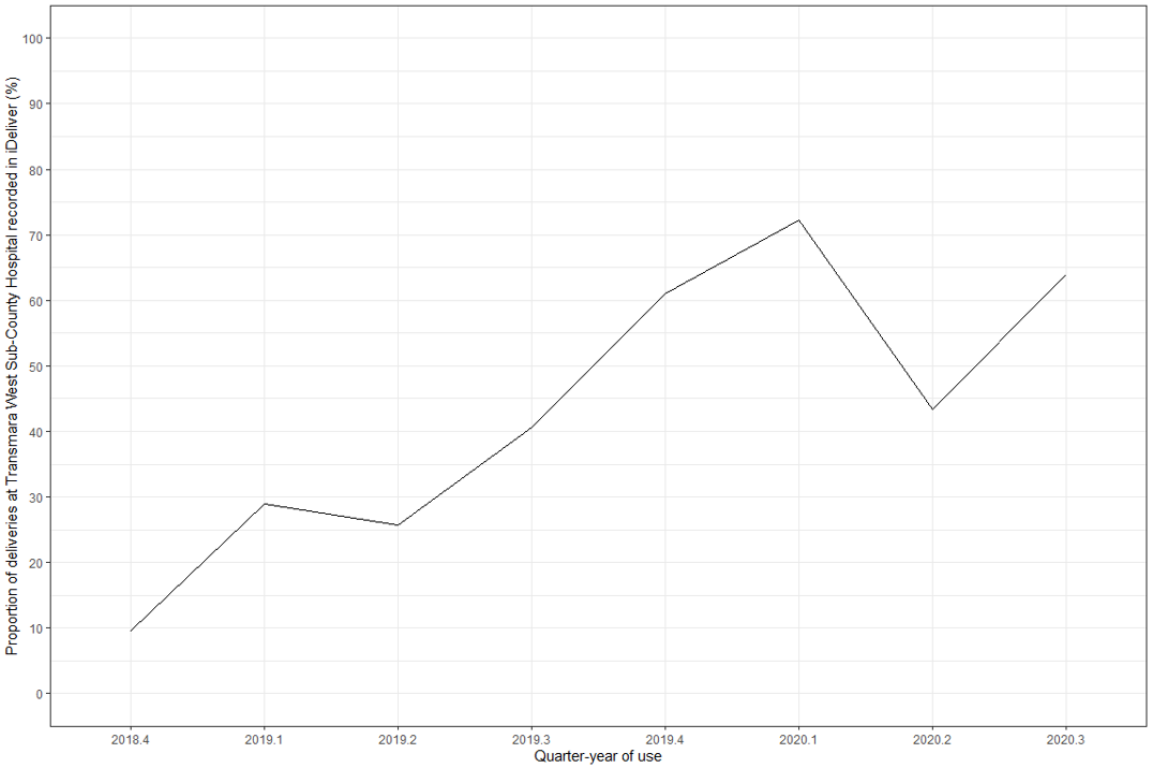
Overview

Data from a total of 1164 deliveries were included in this analysis, spanning 22 months from December 2018 to September 2020. On average, the registered mothers were aged 24.1 years, with a median parity of 1 (range 0-9; Table S1 in Multimedia Appendix 1). Most mothers (996/1164, 85.57%) did not have their education information recorded in iDeliver. Of the 1164 mothers with information recorded, 394 (33.85%) had at least four antenatal visits before their delivery. Table S2 in Multimedia Appendix 2 summarizes the outcomes of the deliveries registered in iDeliver during the study period. Of the 1164 deliveries recorded, most infants were born at term (n=827, 71.05%), with a normal birth weight (n=843, 72.42%), by spontaneous vaginal delivery (n=879, 75.52%). In total, 3.44% (40/1164) of the births were classified as stillbirths (19/40, 48% fresh and 21/40, 52% macerated). Of the 976 babies who were born alive, 943 (96.6%) babies had their vital signs status recorded on discharge—941 (96.4%) were classified as alive and 2 (0.2%) were classified as dead. There was no information on condition at discharge for 3.4% (33/976) babies who were born alive. There was no record of mothers' deaths (0/1164, 0% mothers were classified as dead, and 142/1164, 12.19% of the registration did not have this information recorded).

Assessment of Implementation Outcome: Adoption of iDeliver

Over the 22 months, on average, 45.31% (1164/2569) of the deliveries captured in the existing paper-based record were recorded in iDeliver. The uptake increased by 6.7% on average in each quarter-year, from 9.6% (15/157) in the fourth quarter of 2018 to 64% (235/367) in the third quarter of 2020 (β_1=6.7; SE 2.1; *P*=.005; R^2=0.3). Figure 2 shows the overall increasing trend in the proportion of deliveries recorded by the Transmara West Sub-County Hospital between December 2018 and September 2020, by quarter-year. In the second quarter of 2020, we saw a sharp decline in adoption—a change attributed by local staff to the effect of the COVID-19 epidemic on hospital births. Then, adoption increased again in the third quarter of 2020.

Figure 2. Proportion of deliveries captured in iDeliver compared with paper-based records in the Transmara West Sub-County Hospital, Kenya, by quarter-year from December 2018 to September 2020.



Assessment of Implementation Outcome: Fidelity of iDeliver Use Across the Labor and Delivery Workflow

Figure 3 summarizes the use of iDeliver by each stage of the labor and delivery workflow over the 22 months of data, combined: (1) intake, (2) history, (3) vital signs, (4) labor signs, (5) fetus assessment, (6) quick check, and (7) discharge-labor summary. Of the 1164 deliveries registered in iDeliver from December 2018 to September 2020, the number of data entries was captured in each bar for variables that were organized into these 7 stages. The discharge-labor summary stage of iDeliver had the best data completion rates for all variables within this stage at 72.15% (16,796/23,280), followed by the intake stage (3969/6984, 56.83%), fetus assessment stage (4187/10,476, 39.97%), history stage (6301/18,624, 33.83%), and quick check stage (4945/15,132, 32.68%; Table 1). The data completion rates were lowest for vital signs (1607/9312, 17.26%) and labor signs stages (848/5820, 14.57%; Table 1).

Overall, the completion rate of all variables improved significantly, by 2.9% on average in each quarter-year, from 22.25% (257/1155) in the fourth quarter of 2018 to 49.21% (8905/18,095) in the third quarter of 2020 (β_1=2.9; SE 0.4; P<.001; R^2=0.03). Table 1 also summarizes results from the linear regression analysis assessing the change in data completion of the iDeliver workflow stages over time. Significant increases in data completion in each quarter-year were observed in the history, fetus assessments, quick check, and discharge-labor summary stages (Table 1).

Figure 3. Summary of data completion for all variables at each stage of the iDeliver workflow among all deliveries registered in iDeliver at the Transmara West Sub-County Hospital from December 2018 to September 2020 (N=1164). AMTSL: active management of the third stage of labor; ANC: antenatal care; APGAR: appearance, pulse, grimace, activity, and respiration.

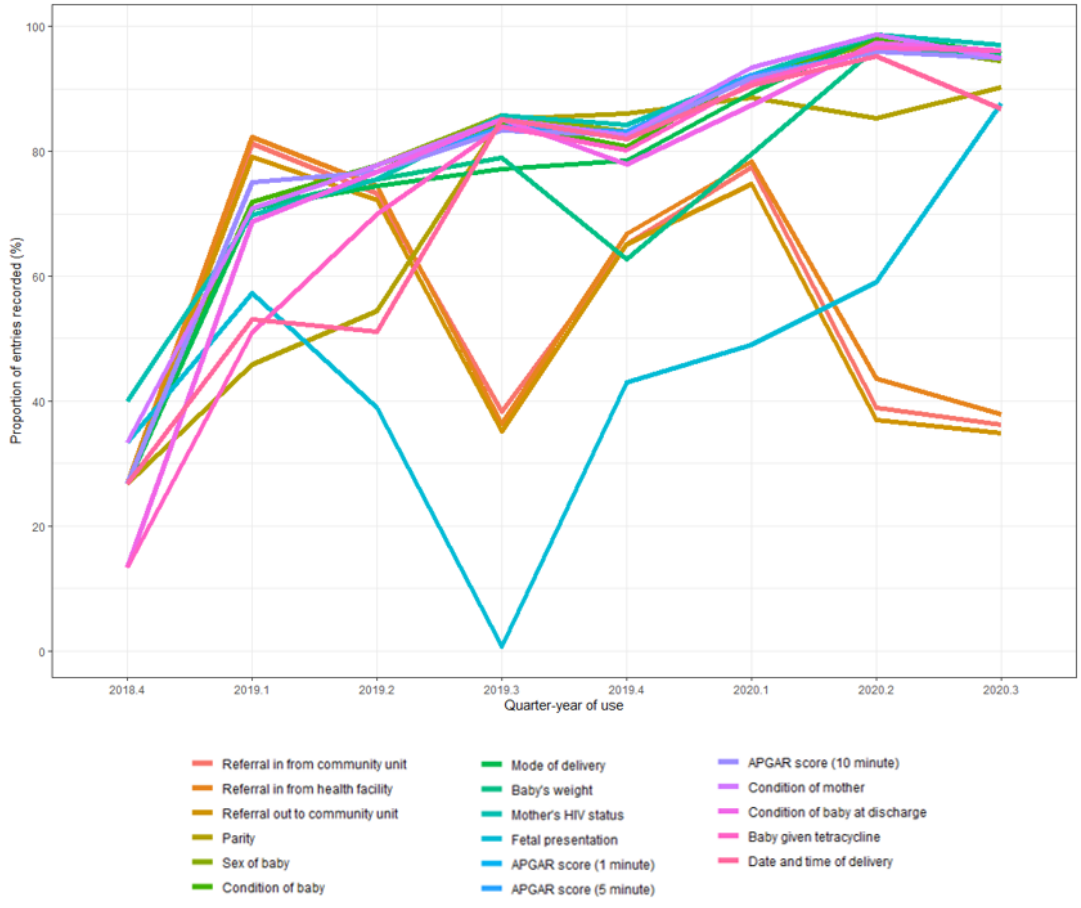
Table 1. Change in the average percentage of data recorded for each stage of the iDeliver workflow at the Transmara West Sub-County Hospital, Kenya, by quarter-year from December 2018 to September 2020.

Labor and delivery stage of the iDeliver workflow	Data recorded from December 2018 to September 2020, %	Change in percentage of data recorded by quarter-year, β_1 (SE)	*P* value
Intake	56.8	−1.2 (1.3)	.40
History	33.8	3.1 (0.9)	<.001
Vital signs	17.3	1.2 (1)	.20
Labor signs	14.6	−1.4 (0.7)	.07
Fetus assessment	40	3.6 (0.8)	<.001
Quick check	32.7	2.6 (0.6)	<.001
Discharge-labor summary	72.1	5.7 (0.7)	<.001
All variables in iDeliver	41.1	2.9 (0.4)	<.001

Assessing the Quality of Indicators From the Kenya MOH's Reproductive and Maternal Health Services

The average data completion rate of the Kenya MOH's key obstetric care indicators increased significantly by 4.6% on average in each quarter-year (β_1=4.6; SE 0.5; *P*<.001; R^2=0.2), from 27.1% (69/255) in the fourth quarter of 2018 to 83.75% (3346/3995) in the third quarter of 2020. The indicators that showed the greatest improvements in completion rate over time were sex of baby; condition of baby at birth; mode of delivery; baby's weight; appearance, pulse, grimace, activity, and respiration scores (at 1, 5, and 10 minutes); baby given tetracycline; condition of baby at discharge; and date and time of delivery—all of these indicators reached completion rate >80% by the third quarter of 2020 (Figure 4). The data completion rate for the fetal presentation indicator showed a steep decrease from the first to the third quarter of 2019 and steadily improved again thereafter (Figure 4).

Figure 4. Proportion of data recorded for indicators that are prioritized by the Kenya Ministry of Health in iDeliver at the Transmara West Sub-County Hospital, Kenya, from November 2018 to September 2020. APGAR: appearance, pulse, grimace, activity, and respiration.



Discussion

Principal Findings

Overview

The data captured by iDeliver indicate that its fidelity and adoption by health workers at the Transmara West Sub-County Hospital tertiary health care facility in rural Kenya showed substantial improvement: use of iDeliver at the Transmara West Sub-County Hospital increased over the 22 months of implementation, by 6.7% on average (SE 2.1) in each quarter-year, from 9.6% (15/157) in the fourth quarter of 2018 to 64% (235/367) in the third quarter of 2020. Data quality also improved over time as the average data completion rate across all variables increased significantly in each quarter-year and for 4 of the 7 stages of iDeliver's workflow. The additional analysis on the data quality of a subset of variables that overlapped with the MOH's maternal and perinatal death notification and review forms also reflected this longitudinal trend of improvement in most variables.

This upward trend in adoption can be attributed to the initial human-centered design efforts, numerous trainings with on-site training on iDeliver updates every 6 months, accounting for staff rotation, full-time on-going support by field staff based in Transmara, and the adaptations made to the iDeliver platform to be responsive to user requests. The uptake of iDeliver, as measured by the proportion of deliveries at the facility that was recorded in the application, decreased briefly in the second quarter of 2020 (Figure 2), reported as the impact of the COVID-19 pandemic, which caused disruption of routine health services including maternal care in Kenya and worldwide [26]. In addition, health care worker strikes occurred in Kenya during this challenging time [27].

A potential reason for the highest data completion rate at the discharge-labor summary stage of iDeliver is that the monthly reports to the District Health Information Software and information for facility billing can be easily generated from iDeliver's discharge-labor summary page. The observed pattern of less iDeliver use during active labor is further explained through an internal evaluation of iDeliver that assessed its usability, acceptability, functionality, effectiveness, and sustainability [28]. Health provider interviews noted that most users entered the data after delivery because they could not input data during a delivery owing to short staffing and that the mother and baby required their full attention until the delivery was complete or *until the gloves come off* [28]. Although this is understandable given the context, it also indicates that the opportunity to use iDeliver for real-time clinical decision-making during delivery is challenging. Patterns of data completeness suggested that users might have preferred iDeliver's feature as a medical record keeping tool and to use it after the active labor and delivery time. On the basis of these findings, the iDeliver team adapted the application to reflect the providers' use culture: a complication management checklist was implemented, modeled after the WHO Surgical Safety Checklist [29], to facilitate a rapid review of the key care principles before or during complication management and the postmanagement documentation. Exploring alternative methods to input data during active labor such as a stylus that can be sterilized, a stylus or tablet inserted into a sterile plastic shield, or dictation capacity can also assist providers with ease of use. Additional sources for clinical guidance have been added to iDeliver so that providers can review global standards of care principles before or after managing a patient with complications, including a web-based version of WHO's Managing Complications in Pregnancy and Childbirth [22], the Merck Manual video library, and the Safe Delivery App [30].

Although the data quality improved over time, completion rates for the vital signs, fetus assessment, labor signs, and quick check stages were still <50%. Checking the vital signs (maternal pulse, blood pressure, and temperature) is important to assess women's well-being during labor and can indicate early signs of complications such as pre-eclampsia, early intrapartum hemorrhage, or impending infection. Examining the progress of labor is necessary to determine whether a woman has inadequate, prolonged, or obstructed labor—all of which can contribute to maternal morbidity and mortality and fetal death. The data missing in the digital system were mostly captured in paper-based records. However, if iDeliver was the only system that captured the data and more than half of the mothers did not have this critical clinical information reported digitally, maternal and fetal welfare could be compromised. Maintaining and improving timely and complete data records in a system such as iDeliver is crucial to capture concrete issues for decision makers to address suboptimal maternal and neonatal outcomes.

Data from iDeliver's aforementioned internal evaluation showed that users valued the benefits of iDeliver for medical record keeping and data storage [28]—a preference also reflected in our analysis, as the labor and delivery summary was completed most by users. Reporting of these indicators also improved significantly over time. Most of this information overlapped with the indicators that we assessed based on the MOH's maternal and perinatal death notification and review. Furthermore, the reported 43 (3.51%) stillbirths out of 1224 deliveries yields a stillbirth ratio of 35:1000 total births—a figure similar to a previous observational study reporting facility-based stillbirth ratio at 38.8:1000 at Kenyan hospitals providing comprehensive emergency obstetric care [31]. These findings provide evidence to support iDeliver's functionality as a platform to provide routine reports (replacing hand-calculated reports) and other key information on quality of care to the facility administrator and the district and national health authority. Users also found that iDeliver's function as an electronic medical record made data extraction for quality improvement audits and health management information systems easier [28].

A use assessment of an intrapartum digital clinical decision support system (CDSS) at the participating facilities in Tanzania and Ghana over 20 months showed that 83% and 67% of all deliveries, respectively, were recorded [32]. In comparison, over 22 months, iDeliver recorded data from 45.31% (1164/2569) of all deliveries overall, with improved uptake from 9.6% (15/157) in the fourth quarter of 2018 to 64% (235/367) in the third quarter of 2020—a figure close to the CDSS assessed in Ghana, in particular. Other studies assessing the implementation of digital CDSSs in Burkina Faso and India

identified common challenges: increased burden for health care providers owing to lack of staff; dual documentation requirements; non–user-friendly platform; high staff turnover; and lack of integration with clinical workflow, continuous training, and staff's motivation [32-34]. On the basis of these insights and the findings from this analysis, strategies to address the implementation challenges facing iDeliver include the following:

Phasing Out Paper-Based Records

The introduction of a digital tool may add to the heavy workload that care providers are carrying, as users need to input data into the new digital system in addition to the paper-based system. This parallel system of double entry was reported as the least-liked aspect of the iDeliver implementation by users, posing the greatest challenge to the application's adoption [28]. To alleviate this burden of work for care providers and improve the adoption, working with facility administrators to phase out paper-based records is a promising strategy. Our experience in Zanzibar supports this premise: a paperless policy was enacted in the major maternity hospital in Zanzibar shortly after iDeliver was implemented there, and birth registrations in iDeliver increased to 100% quickly (Saye, J, unpublished data, April 2022). Furthermore, the significant improvement in data quality of indicators prioritized by the MOH indicates the potential of iDeliver's functionality as a platform to provide key information to facility administrators and replace hand-calculated routine reports to the MOH. Identifying the key indicators, the completion of which must be focused on, such as the MOH indicators, will prioritize data entry and enhance iDeliver's support to quality of care. In October 2020, this strategy was tested by making the MOH indicators mandatory in iDeliver at the Transmara West Sub-County Hospital. The only way for users to not complete these indicators is to skip the whole section that contains the mandatory MOH indicators. The preliminary analysis shows that the average data completeness rate of the MOH indicators increased to 92% after the update. In addition, the adoption of iDeliver also improved, with 88.9% (210/236) of the total deliveries at the Transmara West Sub-County Hospital being captured in iDeliver from October 2020 to April 2021. This improvement further supports the potential to phase out paper-based records to fully adopt iDeliver as the hospital's main medical record system.

Enhancing the Medical Record Function

Insight from both routinely collected data from iDeliver and interviews from users [28] suggest the preference for using iDeliver as an electronic medical record tool. In October 2020, plans were made to implement antenatal and postnatal care modules in iDeliver at the Transmara West Sub-County Hospital, allowing users to keep records of the whole continuum of maternal care for every patient. A dedicated record keeping staff might be considered, but likely difficult to scale-up, given the already widespread human resource challenges.

Enhancing the Clinical Decision Support Function

Even though iDeliver was initially intended as an intrapartum tool, the low coverage of data entry in the vital signs, labor signs, fetal assessments, and quick check stages suggests limited applicability as a clinical decision support tool for intrapartum care in the current context. Additional implementation research is in progress to understand the parameters that health care providers use for clinical decision-making and to derive user-based solutions to adapt iDeliver and improve its usability.

Increasing Motivation for Users

Although the system currently includes user credentials, users at the hospital share devices and often do not log out of their accounts. With this culture of use, we could not assess which type of provider—nurses, midwives, or physicians—used iDeliver the most. Future training of iDeliver should encourage users to log in to the system with their unique credentials so that strengths and areas for improvement for each user can be addressed, which could enhance individual user's accountability. This practice could help identify users who are iDeliver champions—users who use the application diligently and who could train or motivate others. An evaluation of a digital device to reduce maternal mortality and morbidity in low-resource settings reported that the identification and training of key champions, who were clinical staff members who received in-depth training and could support others, was the key implementation strategy enabling the feasibility of that novel intervention [35]. Other evaluation reports from India, Lao, Kenya, and Nigeria also found that staff motivation, satisfaction, confidence, and financial incentives are key factors to enable and sustain the use of novel digital health interventions [34,36,37].

Limitations of This Study

Our study has the strength of leveraging routinely collected data from a newly developed and implemented application over a 22-month period to understand the potential use of an intrapartum decision support tool. However, as this was a retrospective study, these analyses were limited to only what the routinely collected data entailed. This limitation prevented us from conducting a systematic assessment of the behavioral, organizational, and technical determinants [38] of iDeliver's implementation. Additional information from the literature, the field team's insight, and the internal evaluation report were needed to provide insight on how data extracted directly from iDeliver reflected the intervention's adoption and fidelity. In addition, owing to the challenge of data quality, we were not able to adequately assess the process outcomes related to quality of care and any link to maternal and neonatal outcomes. However, data missingness improved substantially over the first 22 months of implementation, and if this improvement persists, we anticipate using these data in the future to assess the impact of iDeliver on the quality of care and on maternal and neonatal outcomes.

Conclusions

The results from this analysis provided us with an understanding of how iDeliver was implemented and used at the facility where it was first introduced. After 22 months, the adoption of iDeliver at the Transmara West Sub-County Hospital showed promising progress, as the use of the application increased and the data quality improved over time. These analyses also suggested that its function as a data collection and reporting tool was used

more than its function as a clinical support tool, triggering the team to promptly adapt the application accordingly to reflect the users' culture of use. This transition has the potential to further improve iDeliver's use, enabling it to timely, correctly, and reliably capture data for both clinical and administrative decision-making support. The iDeliver team will continue to engage with the hospital administrators to support the transition to a paperless workflow and avoid the duplication of workload.

Acknowledgments

This study was supported by Merck and Co, Inc, through the company's MSD for Mothers initiative, known as *Merck for Mothers in the United States and Canada*. The funder had no role in the design of the study or the analyses and interpretations of the results. The authors wish to thank the staff at the Transmara West Sub-County Hospital, Transmara East and West Sub-County Health Management Team, Narok County Health Management Team, and Kenya Ministry of Health for their support during this project. The authors would also like to thank Dr Michael Gravett, Dr Mercedes Bonet, and Dr Zahida Qureshi for their Advisory Committee guidance. For more information, please visit the website [39].

Conflicts of Interest

None declared.

Multimedia Appendix 1
Characteristics of mothers with their deliveries registered in iDeliver at the Transmara West Sub-County Hospital, Kenya, from December 2018 to September 2020.
[DOCX File , 16 KB - formative_v6i6e34741_app1.docx]

Multimedia Appendix 2
Summary of maternal and neonatal outcomes of deliveries registered in iDeliver at the Transmara West Sub-County Hospital, Kenya, from December 2018 to September 2020.
[DOCX File , 16 KB - formative_v6i6e34741_app2.docx]

References

1. Maternal mortality: levels and trends 2000 to 2017. World Health Organization. 2019 Sep 19. URL: https://www.who.int/publications/i/item/9789241516488 [accessed 2022-03-29]
2. Gitobu CM, Gichangi PB, Mwanda WO. The effect of Kenya's free maternal health care policy on the utilization of health facility delivery services and maternal and neonatal mortality in public health facilities. BMC Pregnancy Childbirth 2018 Mar 27;18(1):77 [FREE Full text] [doi: 10.1186/s12884-018-1708-2] [Medline: 29580207]
3. Chuma J, Maina T. Free Maternal Care and Removal of User Fees at Primary-Level Facilities in Kenya: Monitoring the Implementation and Impact: Baseline Report. Health Policy Project. 2013. URL: https://www.healthpolicyproject.com/pubs/400_KenyaUserFeesBaselineReportFINAL.pdf [accessed 2022-04-22]
4. Standards for improving quality of maternal and newborn care in health facilities. World Health Organization. 2016. URL: http://www.who.int/maternal_child_adolescent/documents/improving-maternal-newborn-care-quality/en/ [accessed 2020-04-15]
5. Saving Mothers' Lives: Confidential Inquiry into Maternal Deaths in Kenya: First Report 2017. Ministry of Health, Republic of Kenya. 2017. URL: https://familyhealth.go.ke/wp-content/uploads/2018/02/CEMD-Main-Report-Sept-3-FINAL-Full-Report.pdf [accessed 2020-07-30]
6. Itote EW, Fleming LC, Mallinson RK, Gaffney KF, Jacobsen KH. Knowledge of intrapartum care among obstetric care providers in rural Kenya. Int Health 2019 Jul 01;11(4):258-264. [doi: 10.1093/inthealth/ihy078] [Medline: 30383223]
7. Lusambili A, Wisofschi S, Shumba C, Obure J, Mulama K, Nyaga L, et al. Health care workers' perspectives of the influences of disrespectful maternity care in rural Kenya. Int J Environ Res Public Health 2020 Nov 06;17(21):8218 [FREE Full text] [doi: 10.3390/ijerph17218218] [Medline: 33172160]
8. Bhutta ZA, Das JK, Bahl R, Lawn JE, Salam RA, Paul VK, Lancet Newborn Interventions Review Group, Lancet Every Newborn Study Group. Can available interventions end preventable deaths in mothers, newborn babies, and stillbirths, and at what cost? Lancet 2014 Jul 26;384(9940):347-370. [doi: 10.1016/S0140-6736(14)60792-3] [Medline: 24853604]
9. Campbell OM, Graham WJ, Lancet Maternal Survival Series steering group. Strategies for reducing maternal mortality: getting on with what works. Lancet 2006 Oct 07;368(9543):1284-1299. [doi: 10.1016/S0140-6736(06)69381-1] [Medline: 17027735]
10. Ronsmans C, Graham WJ, Lancet Maternal Survival Series steering group. Maternal mortality: who, when, where, and why. Lancet 2006 Sep 30;368(9542):1189-1200. [doi: 10.1016/S0140-6736(06)69380-X] [Medline: 17011946]
11. Lee AC, Cousens S, Wall SN, Niermeyer S, Darmstadt GL, Carlo WA, et al. Neonatal resuscitation and immediate newborn assessment and stimulation for the prevention of neonatal deaths: a systematic review, meta-analysis and Delphi estimation

of mortality effect. BMC Public Health 2011 Apr 13;11 Suppl 3:S12 [FREE Full text] [doi: 10.1186/1471-2458-11-S3-S12] [Medline: 21501429]
12. Keats EC, Ngugi A, Macharia W, Akseer N, Khaemba EN, Bhatti Z, et al. Progress and priorities for reproductive, maternal, newborn, and child health in Kenya: a countdown to 2015 country case study. Lancet Glob Health 2017 Aug;5(8):e782-e795 [FREE Full text] [doi: 10.1016/S2214-109X(17)30246-2] [Medline: 28716350]
13. Jo Y, Labrique AB, Lefevre AE, Mehl G, Pfaff T, Walker N, et al. Using the lives saved tool (LiST) to model mHealth impact on neonatal survival in resource-limited settings. PLoS One 2014 Jul 11;9(7):e102224 [FREE Full text] [doi: 10.1371/journal.pone.0102224] [Medline: 25014008]
14. Sanghvi H, Mohan D, Litwin L, Bazant E, Gomez P, MacDowell T, et al. Effectiveness of an electronic partogram: a mixed-method, quasi-experimental study among skilled birth attendants in Kenya. Glob Health Sci Pract 2019 Dec 23;7(4):521-539 [FREE Full text] [doi: 10.9745/GHSP-D-19-00195] [Medline: 31874937]
15. WHO guideline: recommendations on digital interventions for health system strengthening. World Health Organization. 2019. URL: http://apps.who.int/iris/bitstream/handle/10665/311941/9789241550505-eng.pdf?ua=1 [accessed 2020-09-20]
16. Monitoring and evaluating digital health interventions: a practical guide to conducting research and assessment. World Health Organization. 2016 Dec 12. URL: https://www.who.int/publications/i/item/9789241511766 [accessed 2021-11-04]
17. Peters DH, Adam T, Alonge O, Agyepong IA, Tran N. Implementation research: what it is and how to do it. BMJ 2013 Nov 20;347:f6753. [doi: 10.1136/bmj.f6753] [Medline: 24259324]
18. Reproductive and Maternal Health Services: Maternal Death Notification Form (MDNF). Nairobi, Kenya: Ministry of Health, Republic of Kenya; Mar 2017.
19. Maternal Death Review Form. Nairobi, Kenya: Ministry of Health, Republic of Kenya; 2016.
20. Reproductive and Maternal Health Services: Perinatal Death Notification Form (PDNF). Nairobi, Kenya: Ministry of Health, Republic of Kenya; Mar 2017.
21. Reproductive and Maternal Health Services: Perinatal Death Review Form (PDRF). Nairobi, Kenya: Ministry of Health, Republic of Kenya; Mar 2017.
22. Managing complications in pregnancy and childbirth: a guide for midwives and doctors. 2nd edittion. World Health Organization. 2017. URL: http://www.who.int/maternal_child_adolescent/documents/managing-complications-pregnancy-childbirth/en/ [accessed 2021-05-03]
23. Oladapo OT, Souza JP, Bohren MA, Tunçalp Ö, Vogel JP, Fawole B, et al. WHO Better Outcomes in Labour Difficulty (BOLD) project: innovating to improve quality of care around the time of childbirth. Reprod Health 2015 May 26;12:48 [FREE Full text] [doi: 10.1186/s12978-015-0027-6] [Medline: 26006170]
24. WHO recommendations: intrapartum care for a positive childbirth experience. World Health Organization. 2018 Feb 7. URL: https://apps.who.int/iris/bitstream/handle/10665/260178/9789241550215-eng.pdf [accessed 2020-08-06]
25. Bartlett L, Avery L, Ponnappan P, Chelangat J, Cheruiyot J, Matthews R, et al. Insights into the design, development and implementation of a novel digital health tool for skilled birth attendants to support quality maternity care in Kenya. Fam Med Community Health 2021 Aug;9(3):e000845 [FREE Full text] [doi: 10.1136/fmch-2020-000845] [Medline: 34344764]
26. Busch-Hallen J, Walters D, Rowe S, Chowdhury A, Arabi M. Impact of COVID-19 on maternal and child health. Lancet Glob Health 2020 Oct;8(10):e1257 [FREE Full text] [doi: 10.1016/S2214-109X(20)30327-2] [Medline: 32758430]
27. As Coronavirus Hits Africa, Doctors and Nurses Prepare to Strike. MEL Magazine. 2020. URL: https://melmagazine.com/en-us/story/as-coronavirus-hits-africa-doctors-and-nurses-prepare-to-strike [accessed 2020-11-18]
28. Avery L. 2019 Internal Evaluation of iDeliver: Key Findings. Winnipeg, Canada: University of Manitoba, Johns Hopkins Bloomberg School of Public Health; 2019.
29. WHO Surgical Safety Checklist. World Health Organization. URL: https://www.who.int/teams/integrated-health-services/patient-safety/research/safe-surgery/tool-and-resources [accessed 2021-02-28]
30. The Safe Delivery App. Maternity Foundation. URL: https://www.maternity.dk/safe-delivery-app/ [accessed 2021-07-22]
31. Aminu M, Bar-Zeev S, White S, Mathai M, van den Broek N. Understanding cause of stillbirth: a prospective observational multi-country study from sub-Saharan Africa. BMC Pregnancy Childbirth 2019 Dec 04;19(1):470 [FREE Full text] [doi: 10.1186/s12884-019-2626-7] [Medline: 31801488]
32. Sukums F, Mensah N, Mpembeni R, Massawe S, Duysburgh E, Williams A, et al. Promising adoption of an electronic clinical decision support system for antenatal and intrapartum care in rural primary healthcare facilities in sub-Saharan Africa: the QUALMAT experience. Int J Med Inform 2015 Sep;84(9):647-657. [doi: 10.1016/j.ijmedinf.2015.05.002] [Medline: 26073076]
33. Zakane SA, Gustafsson LL, Sie A, Tomson G, Loukanova S, Bastholm-Rahmner P. Opportunities and obstacles using a clinical decision support system for maternal care in Burkina Faso. Online J Public Health Inform 2017 Sep 8;9(2):e188 [FREE Full text] [doi: 10.5210/ojphi.v9i2.7905] [Medline: 29026454]
34. Usmanova G, Gresh A, Cohen MA, Kim YM, Srivastava A, Joshi CS, et al. Acceptability and barriers to use of the ASMAN provider-facing electronic platform for peripartum care in public facilities in Madhya Pradesh and Rajasthan, India: a qualitative study using the technology acceptance model-3. Int J Environ Res Public Health 2020 Nov 11;17(22):8333 [FREE Full text] [doi: 10.3390/ijerph17228333] [Medline: 33187163]

35. Vousden N, Lawley E, Nathan HL, Seed PT, Brown A, Muchengwa T, CRADLE Trial Collaborative Group. Evaluation of a novel vital sign device to reduce maternal mortality and morbidity in low-resource settings: a mixed method feasibility study for the CRADLE-3 trial. BMC Pregnancy Childbirth 2018 Apr 27;18(1):115 [FREE Full text] [doi: 10.1186/s12884-018-1737-x] [Medline: 29703254]
36. Numair T, Harrell DT, Huy NT, Nishimoto F, Muthiani Y, Nzou SM, et al. Barriers to the digitization of health information: a qualitative and quantitative study in Kenya and Lao PDR using a cloud-based maternal and child registration system. Int J Environ Res Public Health 2021 Jun 08;18(12):6196 [FREE Full text] [doi: 10.3390/ijerph18126196] [Medline: 34201107]
37. Akeju D, Okusanya B, Okunade K, Ajepe A, Allsop MJ, Ebenso B. Sustainability of the effects and impacts of using digital technology to extend maternal health services to rural and hard-to-reach populations: experience from Southwest Nigeria. Front Glob Womens Health 2022 Feb 8;3:696529 [FREE Full text] [doi: 10.3389/fgwh.2022.696529] [Medline: 35211694]
38. Aqil A, Lippeveld T, Hozumi D. PRISM framework: a paradigm shift for designing, strengthening and evaluating routine health information systems. Health Policy Plan 2009 May;24(3):217-228 [FREE Full text] [doi: 10.1093/heapol/czp010] [Medline: 19304786]
39. iDeliver - VecnaCares. URL: https://www.vecnacares.org/ideliver [accessed 2022-05-23]

Abbreviations

CDSS: clinical decision support system
MOH: Ministry of Health
WHO: World Health Organization

Edited by A Mavragani; submitted 05.11.21; peer-reviewed by T Zan, J Dol; comments to author 25.01.22; revised version received 29.03.22; accepted 31.03.22; published 20.06.22.

Please cite as:
Dinh N, Agarwal S, Avery L, Ponnappan P, Chelangat J, Amendola P, Labrique A, Bartlett L
Implementation Outcomes Assessment of a Digital Clinical Support Tool for Intrapartum Care in Rural Kenya: Observational Analysis
JMIR Form Res 2022;6(6):e34741
URL: https://formative.jmir.org/2022/6/e34741
doi:10.2196/34741
PMID:35723911

Original Paper

Perspectives on Participation in a Feasibility Study on Exercise-Based Cardiac Telerehabilitation After Transcatheter Aortic Valve Implantation: Qualitative Interview Study Among Patients and Health Professionals

Charlotte Brun Thorup[1,2], MSc Nursing, PhD; Anne Villadsen[3]; Jan Jesper Andreasen[1,4], PhD; Jens Aarøe[2], MD; Jane Andreasen[5,6], PhD; Barbara Cristina Brocki[5], PhD

[1]Department of Cardiothoracic Surgery, Aalborg University Hospital, Aalborg, Denmark

[2]Department of Cardiology, Aalborg University Hospital, Aalborg, Denmark

[3]Department of Sociology and Social Work, Aalborg University, Aalborg, Denmark

[4]Clinical Institute, Aalborg University, Aalborg, Denmark

[5]Department of Physiotherapy and Occupational Therapy, Aalborg University Hospital, Aalborg, Denmark

[6]Public Health and Epidemiology Group, Department of Health Science and Technology, Aalborg University, Aalborg, Denmark

Corresponding Author:
Charlotte Brun Thorup, MSc Nursing, PhD
Department of Cardiothoracic Surgery
Aalborg University Hospital
Hoborvej 18-22
Aalborg, 9000
Denmark
Phone: 45 30739950
Email: cbt@rn.dk

Abstract

Background: Aortic valve stenosis affects approximately half of people aged ≥85 years, and the recommended surgical treatment for older patients is transcatheter aortic valve implantation (TAVI). Despite strong evidence for its advantages, low attendance rate in cardiac rehabilitation is observed among patients after TAVI. Cardiac telerehabilitation (CTR) has proven comparable with center-based rehabilitation; however, no study has investigated CTR targeting patients after TAVI. On the basis of participatory design, an exercise-based CTR program (TeleTAVI) was developed, which included a web-based session with a cardiac nurse, a tablet containing an informative website, an activity tracker, and supervised home-based exercise sessions that follow the national recommendations for cardiac rehabilitation.

Objective: This study aims to explore patients' and health professionals' experiences with using health technologies and participating in the exercise-based CTR program, TeleTAVI.

Methods: This study is a part of a feasibility study and will only report patients' and health professionals' experiences of being a part of TeleTAVI. A total of 11 qualitative interviews were conducted using a semistructured interview guide (n=7, 64% patients and n=4, 36% health professionals). Patient interviews were conducted after 8 weeks of participation in TeleTAVI, and interviews with health professionals were conducted after the end of the program. The analysis was conducted as inductive content analysis to create a condensed meaning presented as themes.

Results: Reticence toward using the website was evident with reduced curiosity to explore it, and reduced benefit from using the activity tracker was observed, as the patients' technical competencies were challenged. This was also found when using the tablet for web-based training sessions, leading to patients feeling worried before the training, as they anticipated technical problems. Disadvantages of the TeleTAVI program were technical problems and inability to use hands-on guidance with the patients. However, both physiotherapists and patients reported a feeling of improvement in patients' physical fitness. The home training created a feeling of safety, supported adherence, and made individualization possible, which the patients valued. A good relationship and continuity in the contact with health professionals seemed very important for the patients and affected their positive attitude toward the program.

Conclusions: The home-based nature of the TeleTAVI program seems to provide the opportunity to support individualization, autonomy, independence, and adherence to physical training in addition to improvement in physical capability in older patients. Despite technological challenges, basing the relationship between the health professionals and patients on continuity may be beneficial for patients. Prehabilitation may also be considered, as it may create familiarity toward technology and adherence to the training.

(JMIR Form Res 2022;6(6):e35365) doi:10.2196/35365

KEYWORDS

transcatheter aortic valve implantation; aortic valve; implant; TAVI; telerehabilitation; rehabilitation; aortic stenosis; patients' perspective; older people; elder; aged; geriatric; gerontology; patient experience; user experience; health professional experience; physician experience; telehealth; older adult; telemedicine; cardiac; cardiology; heart; perspective; home-based; exercise; activity tracker; physical activity; mHealth; mobile health; fitness

Introduction

Background

Aortic valve stenosis (AS) has been reported in approximately 8% of octogenarians [1,2]. Living with AS is associated with risk of morbidity and mortality, and the main symptoms related to this disease are weakness, dizziness, syncope, dyspnea, and chest pain when performing daily activities. Decreased independence and quality of life may accompany the symptoms [3]. Ultimately, untreated AS may lead to heart failure and sudden cardiac death [3].

Transcatheter aortic valve implantation (TAVI) is an alternative to surgical aortic valve replacement or medical treatment in patients with symptomatic severe AS. Compared with surgical aortic valve replacement, TAVI is less invasive and is recommended in older patients (aged ≥75 years) or patients at high surgical risk [2,3]. The number of TAVI surgeries is expected to rise over the coming years owing to demographic developments worldwide and the procedure's positive short-term and long-term results [2,4,5].

Cardiac rehabilitation (CR) is recommended after surgery as it improves morbidity, functional capacity, and quality of life [6-8]. Participation in CR after TAVI may be of particular importance because sedentary behavior in this population is related to high risk of mortality and functional decline 1 year after the procedure [9]. International guidelines recommend a multidisciplinary CR approach that includes medical and lifestyle risk factor management, cardioprotective therapies, psychosocial management, exercise training, and health behavior change education to improve functional capacity, recovery, psychosocial well-being, and health-related quality of life of patients with cardiac diseases through risk factor modification [10].

In Denmark, participation in CR after TAVI is low, as <20% of patients are referred to and participate in CR [11]. Old age, lack of availability, and individualized rehabilitation seem to reduce the willingness to participate in CR [10,11].

The use of information and communications technologies (ICTs) in home-based CR, termed as cardiac telerehabilitation (CTR) [12,13], has proved to be comparable with center-based and hospital-based CR programs for mortality, cardiovascular events, cholesterol, blood pressure, BMI, cost-effectiveness, and exercise capacity [14-17].

In addition, CTR may provide an option for patients whose rehabilitation needs are not met by existing services, thus improving attendance rates and adherence, as it is performed in the patients' natural environment and may be incorporated into their daily home routine [17,18].

Supporting older adults in performing exercise at home via a web-based consultation application for a tablet computer was found to be usable [19,20]. In addition, it is expected that different ICTs may be used advantageously to collect and transfer data from the patient to a digital platform or a personal health record in addition to web-based consultation devices for training and communication. However, we have not found any studies that focus on CTR that are based on physical activity and target patients after TAVI surgery. Whether CTR in this population may improve patients' physical functioning, recovery, psychosocial well-being, and quality of life is unknown. The use of ICTs to improve enrollment and adherence to rehabilitation and support medical and lifestyle risk factor management after TAVI has not been investigated. In addition, it is unknown how patients and health professionals experience participation in CTR. Thus, this perspective requires further investigation to create tailored CTR programs for this target group. On the basis of a participatory design [21], we developed a 12-week digital CTR program, named TeleTAVI, which is ready to be tested in a feasibility study.

Aim

This study was a part of the feasibility study and will only report on the qualitative findings; therefore, the aim was to explore patients' and health professionals' experiences with using health technologies and participating in the exercise-based CTR program, TeleTAVI.

Methods

Design

This study reports the qualitative findings from the feasibility study and follows the Consolidated Criteria for Reporting Qualitative Research guideline [22].

Ethics Approval

Owing to the qualitative nature of the study, the Regional Ethics Committee stated that no approval was required. According to the Helsinki Declaration, oral and written information was provided and informed written consent was obtained from all

patients and health professionals. The study protocol was approved by the head of the department and registered by the hospital (ID 2020-054).

Recruitment

Patients were recruited from the Department of Cardiology, Aalborg University Hospital, Denmark, between March 8, 2021, and May 25, 2021. The TAVI surgery was performed under conscious sedation, and the patients returned to the ward on the same evening or the next morning and were discharged within 2 to 3 days after the surgery.

Inclusion and Exclusion Criteria

All participants in the qualitative study were the patients and health professionals participating in the feasibility study of TeleTAVI. Inclusion criteria for the patients in the feasibility study were being adults (aged ≥18 years) who underwent a TAVI surgery and could read and understand Danish. The patients in the present cohort were primarily older people with high-risk symptomatic AS. Patients with physical deficits adversely influencing physical performance, as measured by the 6-minute walk test [23], and patients with decreased cognitive functioning, as assessed by the Mini Mental Scale evaluation, were excluded [24]. We also excluded patients with no internet access or low data coverage at their home address. Eligible patients were approached for inclusion on the day before surgery. All patients with cardiac diseases discharged from a Danish hospital can participate in municipality-based CR after the first 8 weeks. This was not an exclusion criterion, but none chose to participate in CR parallel to web-based training. Inclusion criteria for health professionals were having experiences of care, treatment, and rehabilitation targeting older patients with cardiac diseases receiving TAVI.

Telerehabilitation Technologies and Intervention

TeleTAVI was created as a 12-week digital CTR program that included supervised home-based video training, a web-based session with a cardiac nurse specialist, an informative website, and an activity tracker (Textbox 1) to be used during 8 of the 12 weeks.

Textbox 1. Technologies used in the TeleTAVI program.

Tablet (Apple iPad; Wi-Fi; 10.2 inches; 4G)

- Outlook mail program (Microsoft 365 Office) for web-based exercise training and communication with project personnel via a videoconferencing system, *Videosamtale application*, installed in the unit. The system allowed the physiotherapist to see all the patients on the same screen and communicate with the patients. In addition, patients can see and hear the physiotherapist at the hospital, see themselves and other patients on their screen, and hear all conversations.
- Access to the project´s website
- Visual access to a personal record with information and graphs on uploaded data on daily steps

Website (Multimedia Appendix 1) hosted by the North Jutland region and installed in the tablet

- Text information on issues regarding treatment, lifestyle, and medicine
- Videos with training programs
- Videos with patients' experiences and information from health professionals

Activity tracker worn on the wrist during daytime (Beurer Activity Sensor AS97; with the use of Beurer AS 97, it was possible to connect and store data on the information technology system used at the hospital, thus complying with the General Data Protection Regulations Compliance Guidelines in Europe)

- Tracks daily step counts and heart rate during the supervised training sessions.
- Patients were provided an alternative to either upload data on number of daily steps to the personal record or register their daily steps on a personal diary at the end of each day.

Booklet in paper form

- Schedule for home visits, web-based training, and self-training and for charging the iPad and the activity tracker
- User manuals for the tablet, website, activity tracker, and Outlook
- Description of home exercises, also illustrated by pictures

Training equipment used by patients during web-based training sessions

- Step bench, training mat, elastic exercise band, and dumbbells (1, 2, and 3 kg)

A display on the activity tracker showed the number of daily steps taken, and through uploads, the steps were also visible in the patients' personal records on the tablet. The activity tracker also had a heart monitor to determine the intensity of the training. In addition, the program included a booklet containing schedules for home visits, web-based training, and self-training; user manuals for the tablet, website, activity tracker, and Outlook mail program; and a description of home exercises illustrated by pictures and written instructions. The website included videos of patients and relatives presenting experiences with TAVI treatment in addition to video-based training programs to be used for unsupervised supplementary training

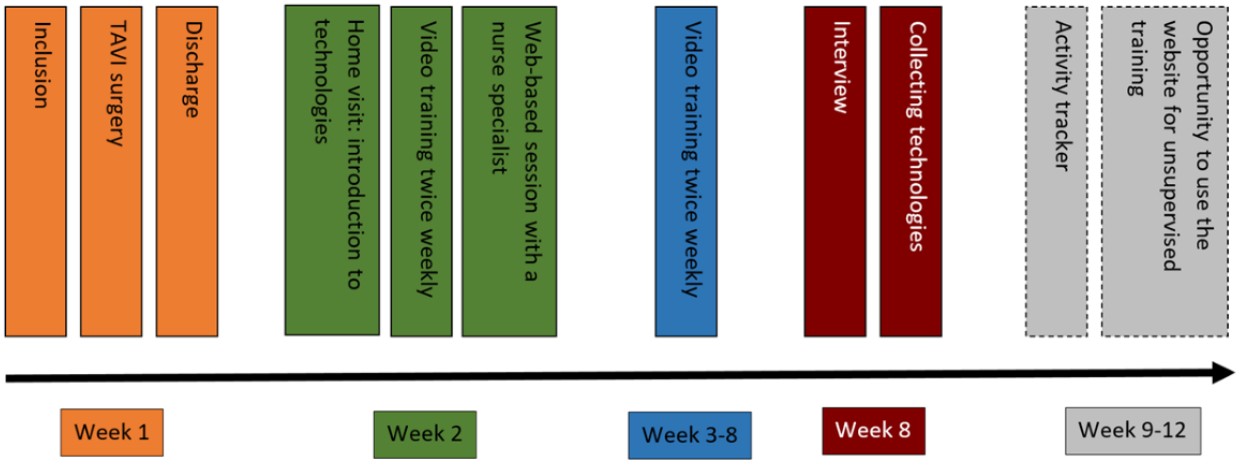
during the first 8 weeks and after, if the patients have their own tablet or computer at home (Multimedia Appendix 1). The patients could choose to maintain the activity tracker for the entire 12-week study period.

The technologies used in TeleTAVI were delivered to the patients in their homes 1 week after hospital discharge. Before hospital discharge, the patients were instructed on a short strength training program for home exercising to be used until the telerehabilitation technologies were delivered to the patient.

To ensure safety during the sessions, the patients were introduced to the exercises and technology by the last author (BCB) during the first home visit. If needed, additional technical support was provided through home visits or telephone calls. The home exercise training was individualized, based on the physical functioning as assessed before surgery and during the first home visit by patients trying the different exercises and equipment to be used in the web-based training sessions. Individualized goals for training were set and followed the national recommendations for CR, with a combination of aerobic and strength training twice weekly, with each session lasting from 30 to 60 minutes [25]. In addition, patients were instructed to take a 30-minute walk daily at moderate intensity.

The TeleTAVI program consisted of web-based physiotherapist-supervised exercise training in groups, twice weekly via a tablet; an activity tracker (Textbox 1; Figure 1) that collected the daily number of steps and was able to monitor heart rate; a web-based session with a cardiac nurse specialist; a textbook; and an informative website (Multimedia Appendix 1).

Figure 1. Timeline for TeleTAVI. TAVI: transcatheter aortic valve implantation.



Qualitative Data Collection and Analyses

The rehabilitation period was 12 weeks, and the duration of the supervised web-based training was 8 weeks, starting in the first week after surgery. After the initial 8 weeks, patients were instructed to continue to exercise unsupervised, and the interviews were conducted at the same time. This time was chosen to gain comprehensive insight into the web-based rehabilitation period without any recall bias.

Individual interviews were conducted with all the patients who had completed the intervention. Furthermore, individual or group interviews were conducted with the involved health professionals. All interviews were based on a semistructured interview guide [26], covering issues such as patients' experiences with video training, web-based session with a nurse, usability of technology and website, patient support, and issues that the patients felt were important. For health professionals, the focus of the interviews was on their experiences with web-based training and communication. CT (first author) and BB conducted interviews at the patients' homes, and if present, relatives were encouraged to participate. For the health professionals, the interviews were conducted at the hospital after a minimum of 8 weeks of experience with TeleTAVI for each health professional. These were conducted either as group interviews (physiotherapists) or individual interviews (nurses) for practical reasons. Both interviewers had previous experiences with conducting qualitative interviews. The analysis was conducted as inductive content analysis to create a condensed meaning presented as themes [27]. The recorded interviews were transcribed verbatim and read several times to familiarize with the text and obtain an overall sense of the data. Then, the data were coded for manifest and latent content to identify themes that elaborated the underlying meaning using the coding system, NVivo (version 8; QSR International) [28]. The analyses and interpretations were primarily performed by CBT, AV (second author), and BCB based on their agreement that essential themes were reached [29]; then, the findings were discussed in the research group.

Results

Overview

During the study period, 41 patients were referred to TAVI and screened for eligibility. Of these 41 patients, 13 (32%) patients declined to participate, and 13 (32%) patients lived in a place with no internet connection or poor data coverage, thus failing to meet the inclusion criteria. Of the 37% (15/41) of the patients enrolled in the study, 47% (7/15) withdrew after the surgery owing to tiredness (2/7, 29%), non–cardiac-related hospital readmission (2/7, 29%), fluctuating health (1/7, 14%), and regretting participation when the health technologies were introduced during the home visit (2/7, 29%). Furthermore, 7%

(1/15) of the patients died before hospital discharge. All patients (7/7, 100%) who completed the TeleTAVI program were interviewed (Table 1). In addition, 2 nurses and 2 physiotherapists (Table 2) who participated in the telerehabilitation program were interviewed. The interviews lasted between 35 and 50 minutes for patients, between 15 and 25 minutes for nurses, and 55 minutes for the group with the physiotherapists.

The patients were aged between 74 and 90 years and had different experience levels in using a tablet or computer. Patients with no experience were able to receive guidance from their relatives, making it possible to complete the telerehabilitation program. In addition, the patients had different physical capability levels after the TAVI procedure, as some felt limited in performing sport and exercise. Finally, some of the patients had comorbidities that may have affected their physical capability, and 57% (4/7) of them had Tilburg frailty score ≥5, indicating frailty in these patients. Regarding education level, 14% (1/7) of the patients had primary school education, 71% (5/7) had vocational education, and 14% (1/7) had higher education (Table 1). The health professionals were aged between 25 and 42 years and had work experience ranging from 2 to 13 years (Table 2).

The analysis and interpretation of the patients' and health professionals' interviews resulted in 3 themes and 9 subthemes (Textbox 2).

The interviews revealed that patients and health professionals experienced diverse technological issues. All patients expressed interest to participate in the TeleTAVI rehabilitation program, but none of them were inspired or motivated by the health technologies used. Despite being challenged when using the technologies, they were all optimistic about participation.

Table 1. Characteristics of the interviewed patients (N=7).

ID	Sex	Age[a] (years)	IT experience	Living alone	Education level	Tilburg frailty score[b,c]	Comorbidities	Limitation in doing sport or exercise
P[d]1	Male	85	PC	No	Vocational education[e]	2	AF[f]	Low
P2	Female	90	Tablet	No	Higher education	9	IHD[g] and AH[h]	Moderate
P3	Female	84	Tablet	Yes	Primary school	2	AH	None
P4	Male	87	None	Yes	Vocational education	8	AF and IHD	High
P5	Male	82	Tablet and PC (spouse)	No	Vocational education	6	IHD	None
P6	Female	74	Tablet	Yes	Vocational education	4	AH	Moderate
P7	Female	81	Tablet and PC (spouse)	No	Vocational education	9	None	Moderate

[a]Median: 84 (range 74-90) years.

[b]Median: 6 (range 2-9).

[c]Score ≥5 points is considered as frailty when using the Tilburg frailty score [30,31].

[d]P: patient.

[e]Primary school and 2 to 5 years of vocational education.

[f]AF: atrial fibrillation.

[g]IHD: ischemic heart disease.

[h]AH: arterial hypertension.

Table 2. Characteristics of the health professionals (N=4).

ID	Sex	Age[a] (years)	Profession	Work experience[b] (years)
HP[c]1	Female	33	Nurse	5
HP2	Female	42	Nurse	13
HP3	Male	34	Physiotherapist	3
HP4	Male	25	Physiotherapist	2

[a]Range: 25-42 years.

[b]Range: 2-13 years.

[c]HP: health professional.

Textbox 2. Themes and subthemes derived from the content analysis.

Technological challenges

- Reticence toward using the website
- High interest and low competence regarding the activity tracker
- Anticipating technical problems led to worries beforehand.

Advantages of home-based training

- Individualization despite lack of hands-on guidance and technical problems
- Home-based training reduces transportation time and may support adherence.
- Proper exercise was done.

The importance of establishing a relationship

- Web-based training does not support relatedness between patients.
- First visit clears the way and creates a base for continuity.
- No recall of the web-based session with the nurse.

Technological Challenges

Reticence Toward Using the Website

Most of the patients (6/7, 86%) did not visit the website. Some were not curious about the content, whereas others were not familiar with the technologies and were afraid of making errors, such as deleting elements unintentionally. A participant described the following:

> *Actually, we haven't [visited the website]. We were afraid of pressing a key that would delete some of the content.* [P1]

The participants' concern about making errors limited the use of the website, and therefore, they did not fully benefit from the knowledge and coping elements provided on the website. Therefore, the patients' overall benefit from the website may be questioned.

Other reasons for not visiting the website were reluctance toward being involved in other patients' experiences of the disease:

> *I've always felt like this...I can't stand to hear about disease, about faults in the heart.* [P2]

This finding shows that some elements on the website may have unintentional consequences by triggering additional anxiety and fear when confronted with other patients' experiences of living with heart disease.

Furthermore, few patients (2/7, 29%) watched the training videos, and only 14% (1/7) of the patients performed additional workouts. They argued that training twice a week with the physiotherapist felt sufficient.

Overall, the participants had neither the skills nor the curiosity to explore the content of the tablet, meaning that they did not make full use of the opportunities given. Others did not want to be confronted with other patients' life stories.

High Interest and Low Competence Regarding the Activity Tracker

All patients (7/7, 100%) found the activity tracker exciting to use, as they could keep track of how many steps they had walked, and the tracker made it possible to compare steps from day to day, which facilitated motivation. All but 1 patient (6/7, 86%) preferred to write down the daily steps in a diary, as it was very difficult to upload daily steps via Bluetooth to their health record. In addition, some patients missed the information that the tracker had to be recharged once a week, leading to additionally missing electronically uploaded step data. Furthermore, the monitoring of heart rate during exercise was difficult for the patients to learn, meaning that the intended use of heart rate to determine exercise intensity was lost:

> *Well, you think more about it [improvements]...and you are excitedly waiting...Oh, today you have walked this much.* [P7]

> *That's the only thing I have used it for, measuring steps. I have not measured heart rate...well, I think, I have learned to measure steps, lets stick to that. Because I am not fond of learning new things...I would rather avoid it.* [P2]

Although the activity tracker motivated walking activities, the participants were challenged technically. Thus, the potential motivational factor for the patients to follow their increase or decrease in daily steps walked was hampered, and the health professional's opportunity to continuously follow the patients' walking activity was lost for most of the patients (5/7, 71%).

Anticipating Technical Problems Led to Worries Beforehand

Some patients described preparation for sessions as slightly compulsive and held themselves in readiness twice a week, worrying that technical problems would occur, such as failing to log in or malfunctioning sound and vision in the tablet. The patients described the technical problems during training as problems with hearing the physiotherapist clearly, stuttering

web-based transmission, and feeling interrupted by other patients:

> *It was scaring, I was worried for the web-based sessions. In good time I was sitting in front of the tablet, wondering whether it would work this time or not. Sometimes I couldn't hear or see anything, or they could not hear me.* [P2]

Thus, the patients reported that waiting for the web-based sessions twice a week felt slightly involuntary and created worries that something would go wrong.

Advantages of Home-Based Training

Individualization Despite the Lack of Hands-on Guidance and Technical Problems

The physiotherapists reported technical problems during the web-based training sessions, such as problems with logging in and handling the sound and picture settings. Delays in starting the session and interruptions during training were frequent. This led to 2 physiotherapists being present during sessions, one managing technological problems and the other conducting the training sessions:

> *Sometimes, they needed a lot of guidance because...then they switched the camera on in the wrong direction, or the screen did "freeze" or, they just happened to turn it off. You had to guide the patients through all this, and when the training had begun, another physiotherapist guided the patient through technical problems.* [HP3]

The physiotherapists questioned the quality of the exercises performed, as the web-based training made it difficult for them to guide the patients in the usual hands-on way. In addition, intermittent poor internet connections, stuttering images, and very small images made it difficult for the physiotherapists to see whether the exercises were performed correctly and properly by the patients.

The physiotherapists observed great diversity in the patients' physical and respiratory fitness, thus spending more time on patients who were most challenged. However, they strived to achieve individualization, to support all patients in gaining improvements in fitness level despite individual capabilities. For the physiotherapists, the individualization seemed double-sided as it was time-consuming; however, it created a possibility for all patients to improve their level of physical capability. On the basis of these experiences, the physiotherapists suggested that in future interventions, patients should be divided into groups based on their level of physical capacity.

Similarly, the patients valued that the training was individualized and appreciated being guided in the correct execution of the exercises:

> *And when there is something you can't do, then he [the physiotherapist] always corrects you in a good way.* [P5]

This meant that the physiotherapists had to accept that the quality of the training might be reduced, as hands-on guidance was not possible together with the patients' limited technical skills and slow internet connections. However, the patients expressed that the opportunity for individualization supported improvement.

Home-Based Training Reduces Transportation Time and May Support Adherence

Most of the patients (5/7, 71%) increased their web-based technical competencies, whereas others needed help throughout the study period. Despite the web-based technical problems, both patients and physiotherapists provided several positive feedbacks about the home-based web-based training. All patients (7/7, 100%) appreciated being able to exercise in their homes because it felt safe, and they avoided transportation to a rehabilitation center, which was one of the main reasons for participating in TeleTAVI for all patients:

> *I did benefit from participation, also in a physical way, by experiencing that nothing fatal would happen during exercise...I did manage it without collapsing halfway through.* [P1]

> *I mean, it's been easy because you were at home, and you didn't have to drive for it...and the time [was used properly].* [P7]

In addition, some patients felt inspired to continue web-based training after 8 weeks of participation. This was consistent with thoughts expressed by physiotherapists, who stated that home-based exercise might motivate patients to continue training after the project has ended because the patients were taught in their homes and felt safe while performing the exercises there:

> *Well, we can just keep on doing the exercises...we will continue to do that.* [P5]

> *The advantage is that it's easily transferable for them [the patients] afterward.* [HP3]

Thus, avoiding the costs and time spent on transportation and the expectation that home-based training may support adherence to future training activity were positive aspects of participation.

Proper Exercise Was Performed

All but 1 of the patients (6/7, 86%) felt physical improvement at the time of the interviews. Whether improvements were mainly caused by the TAVI surgery, telerehabilitation, or most likely, a mixture of both cannot be determined in this study.

Despite considerations about own physical ability, approximately all patients felt that they had performed a proper and satisfying exercise for the whole body, leading to a natural feeling of tiredness after the sessions. Some participants became more energized and apprehended that less dyspnea made them capable of exercising at a high level:

> *Well, I can feel that it's good for my body. It's like, I become livelier and light. And I think that you gain energy when you do exercise...when I am finished with workout, my hair is wet, and my clothes needs washing.* [P6]

> *I have realized that exercise helps you, and now...as I am less dyspnea, more comfortable and my feet got*

> *smaller. I can actually be more active and not just "dragging around."* [P4]

Being under surveillance and encouraged to challenge one's comfort zone during exercise seemed to increase the benefit of training. The patients' old age was kept in mind while considering their expectations of improvement, realizing that there must be a limit on how much improvement to expect. In contrast, a patient expressed that exercise may "stop the clock" and add extra time to their life:

> *I think, when you are forced to do more than you actually can, then it pays back...and I do more during training than I would have done on my own...You have to remember that my body is 90 years of age; it is limited how much better I can become.* [P2]

> *You got sweatier exercising with the physiotherapist than when you exercise on your own... [when becoming older] training might stop the clock.* [P1]

On the basis of the physiotherapists' observation of patient improvements, they agreed that the web-based sessions fulfilled the goals for exercise training, such as improvements in fitness level and muscle strength, and increased the feeling of being safe during training:

> *It succeeded because their general functional level and increased age do "set them back" [physically], meaning that just small efforts provide minor improvements; even though the patients don't do the exercise correctly, there might still be physical progress.* [HP3]

Thus, patients and physiotherapists felt that exercise improved physical fitness, and both surveillance and encouraging pressures seemed to support the patients' feeling of exercising safely.

The Importance of Establishing a Relationship

Web-Based Training Does Not Support Relatedness Between Patients

The physiotherapists expressed worries that the TeleTAVI would reduce patients' opportunity of being socially involved with each other. For the patients, being socially involved with others seemed less important and did not influence their willingness to participate in the program. Being able to see others during training did not have any importance for most patients (6/7, 86%), as they were just considered as "images on the screen" and not as living persons. Others were only noticed if they disturbed the sessions. When relatives participated in the exercise with the patients, they performed the exercise outside the reach of the camera:

> *The most annoying about the tablet, is that you are forced to look at the others while training. The others don't matter to me because I don't look at them as persons [...]. And one, she was talking so loudly...I couldn't understand what the turmoil was about [...]. I was very close to just turning the s*** off; I couldn't deal with all that commotion.* [P6]

As such, being socially related to other patients was not considered as an important issue by the patients.

First Visit Clears the Way and Creates a Base for Continuity

All patients (7/7, 100%) had the training equipment and tablets delivered at home. Being introduced to the program and having the opportunity to try both the technologies and training equipment was beneficial in preparing the patients for the following web-based sessions. Some patients felt excited, whereas others expressed that meeting the physiotherapists was nice and helped them in performing the exercise safely and properly:

> *It was really nice [having the technology and equipment brought to the home], we had a really good talk, and I was happy to be shown how to do the exercises properly.* [P1]

The relationship established between BCB and the patients during the first home visit and through additional visits became important, and they felt grateful. In particular, physiotherapists' ability to be attentive toward each individual patient during the web-based exercise sessions made the patients feel that they could participate on equal terms. Continuity in the relationship with the health professionals was expressed as very important. The patients felt acknowledged as individuals, and thus safe.

Thus, it was evident that the first home visit affected the relationship with the patients positively and supported a good start within the program. In addition, seeing the same person at the first home visit and several times during the rehabilitation period positively influenced the patients' attitudes toward the program.

No Recall of the Web-Based Session With the Nurse

During the nurse session, 1 week after discharge, both nurses experienced difficulties with the video call, meaning that some of the sessions were performed as telephone calls. Both nurses expressed that video calls had many advantages, such as being able to see the patients' physical condition and obtain a sense of their mental and psychological well-being. The video call provided security and continuity for the patients, as they may have met the nurse during hospitalization:

> *They were happy about it [the web-based session]. I think for them it felt comforting that someone did follow up on them about these things [health and well-being after discharge].* [HP2]

Most patients (6/7, 86%) did not remember talking to a nurse, and if they remembered, they had little or no recollection of what was discussed:

> *A nurse? ...I think it was on the phone...no, wait, it might have been one from the healthcare center, sorry, I can't tell. And I can't remember what we talked about.* [P3]

Thus, despite good experiences from the nurses' point of view, none of the patients could recall the session's content, and most of them (6/7, 86%) did not remember the session at all. An explanation may be that this group of older people have numerous contacts with health professionals, either from the hospital or the primary sector, making it difficult to discriminate between the health professionals. However, it seems to underpin

the importance of supporting continuity in the contact with the patients and to question the timing and content of the call.

Discussion

Principal Findings

This study aimed to explore patients' and health professionals' experiences with using health technologies and being part of the exercise-based CTR program, TeleTAVI.

In summary, the interviews indicated that patients were slightly reticent toward using the website, they had neither the skills nor the curiosity to explore the tablet, and some avoided other patients' life stories. The activity tracker held much interest from the patients, but they did not gain full benefit from it as their technical competencies were low. This was also a problem when using the tablet for web-based training sessions, leading to patients feeling worried before the training, as they anticipated technical problems. The physiotherapists expressed that the disadvantages were technical problems and not being able to use hands-on guidance when the patients' exercises were performed inappropriately or incorrectly. In contrast, they expressed that improvements were observed in patients' physical fitness and that training at home created a feeling of safety and supported adherence to training. In addition, they avoided transportation to a training center. Despite the low quality of pictures and, sometimes, the internet, the physiotherapists made individualization possible. Individualization was valued by the patients, and the they felt that they exercised properly and felt improvement during the rehabilitation period. Good relationships and continuity in contact were extremely important for the patients and affected their attitude toward the program; when having only one contact, as the nurses did, the recall of the session was hampered. None of the patients (0/7, 0%) expressed a wish to be socially related to other participants during the program.

Comparison With Previous Work

Technological Challenges

Technological challenges were evident when talking to the health professionals. These findings are consistent with those of a review by Fischer et al [32] on the acceptance and use of health technology in older people. They concluded that older people face different challenges, such as limited familiarity with technology, reticence about asking for assistance, mistrust, and concerns about privacy when using technology [32]. This was similar to the patients in this study who were reticent toward using the technology and had neither the skills nor the curiosity regarding the use of the tablet and website. In addition, they were afraid of deleting something from the tablet, which refrained them from exploring possibilities it.

Older patients' preferences and attitudes toward digital technology have been investigated by Terp et al [33], who concluded that lack of knowledge, user competence, and interests were the main barriers to older people's use of technology. These findings are consistent with the results of this study, in which the patients did not use the full potential of the tablet, activity tracker, or heart monitor and none of them participated owing to interest in using health technologies. This underpins that devices targeting older people should be easy to use and recharge, with automatic uploading of data.

Another explanation for the low use of technology may be related to their level of education. Hargittai et al [34] studied internet skills among citizens aged ≥60 years and concluded that high income and high level of education equals high level of internet skills. This contributes to the explanation of why patients in this study experienced problems in using the tablet and activity tracker. It is interesting that not being interested in or not having experience with technology did not discourage them from participation; in contrast, they were glad to do so, which indicates that they were not afraid of trying new and unknown technology. This is consistent with Fischer et al [32], who concluded that older people generally have a positive attitude toward using health technologies at their homes, if they believe that it will help the health professionals to preserve their independence and autonomy. In addition, they seem less bothered when technical problems arise; they just wait for problems to be solved [33]. In this study, the patients did not express any annoyance when facing technological issues. They got used to 2 physiotherapists being present during the exercise sessions, one helping with technologies and the other guiding the training, and they did so without questioning this high use of health personnel resources.

Thus, mistrust toward technology and reduced familiarity, competencies, curiosity, and trust in one's own ability were important factors that influenced the patients' willingness and possibility of using the technology, resulting in not gaining full benefit from the technology used. In contrast, this did not discourage them from participating in the CTR program in which the training could be conducted in their home environment.

Home-Based Training and Establishing a Relationship

The patients were satisfied with avoiding transportation and being able to exercise at home. They felt having performed proper exercise during the web-based sessions and felt improvement in health, physical status, and life in general. However, life expectancy and quality of life may not mean the same across generations. Leeuwen et al [35] found that, for older adults, quality of life is related to autonomy, independence, and ability to handle life circumstances and potential changes that come with increased age. In addition, experiences of one's own health seem relative as experiences with health depend on circumstances and what a person finds to be reasonable in relation to their age, history, medical condition, and social situation [35]. A study on older men's perspectives on good and healthy aging reveals that becoming older is balancing having expectations and ambitions for your life and realizing the realities about your physical and social situation. This balance may be handled either by improving your circumstances in life or lowering your ambitions and expectations [36]. They concluded that physical and cognitive health was important to ensure independence and autonomy [36]. In this study, the patients felt improvement, whereas the physiotherapists questioned the quality of exercise; however, it might be argued that life expectancy and quality of life in older people may relate

more to maintaining independence and autonomy than an increased physical health level. Thus, their expectations for participating in telerehabilitation may be the improvement of life circumstances, while at the same time, accepting their everyday life capabilities.

Social interaction and connectivity to peer fellows have previously been considered as important in supporting behavior changes, such as increasing physical activity [37,38]. However, it is argued that some commonly used self-regulation intervention techniques that are effective for younger adults may not be effective for older adults [38]. This is consistent with our study, in which none of the patients expressed a wish to be socially related to other participants during the program.

Professor in Sociology, Arthur Frank [39], describes that one way of experiencing and talking about illness is to rely on a restitution narrative that holds the expectation that a sick body will become better. All the patients in TeleTAVI participated owing to expectations of becoming better and recovering, and they trusted that telerehabilitation would support this. Thus, it can be argued that as the patients' narratives follow this restitution model [39], it is apparent that the patients feel motivated to participate in a rehabilitation program such as TeleTAVI because their main goal is to recover after their TAVI surgery. This could mean that there is great incentive to continue to develop telerehabilitation for older patients in general because, from the patients' perspectives, there are several strengths to be found in telerehabilitation.

In summary, older people may view life expectancy and quality of life as a balance between life expectations and circumstances regarding age, physical and medical condition, and social situation, and motivation for participation in CTR may originate from a restitution narrative. It can be argued that this group of patients might be the ideal target group for telerehabilitation because their advanced age and decreased activity reduce their willingness to participate in rehabilitation outside their homes and individualized exercise performed in their own homes may increase independence, autonomy, and adherence. Hence, the main challenge for telerehabilitation is to restructure and further develop it to make it more efficient and feasible, which will be discussed in the following section.

Future Directions

One of the main strengths of telerehabilitation that both the health professionals and the patients emphasized was the possibility to individualize the web-based exercise sessions. Eichler et al [40] argued that individualization of CR after TAVI is important to maintain the patients' autonomy.

A study by Eichler et al [40] about geriatric versus CR concluded that it may be rewarding to classify the degree of frailty regarding older patients' technological and physical skills. Dividing patients into different groups would ensure that those who need more support receive it. This was consistent with the suggestions from the physiotherapists in this study.

In addition, individualization and enhanced efficacy may be reached by starting rehabilitation before the TAVI surgery as prehabilitation. Pighi et al [41] investigated the determinants of outcomes after TAVI and found that high frailty before surgery negatively affects the outcome of the surgery, particularly in women. Therefore, Pighi et al [41] suggested the implementation of prehabilitation to support the physical status of patients who are frail and to have better outcome after TAVI [41]. Prehabilitation also holds the opportunity to practice their technological skills before the surgery.

In this study, training at home supported the patients in feeling safe, and it was expected to enhance adherence to exercise after the rehabilitation period. The long-term effects of exercise training after TAVI have been investigated by Pressler et al [42], who argued that it is important to continue exercising to maintain improvements in the long term, but most fail to adhere to exercise [42].

Thus, the feasibility of TeleTAVI in the current form may still be questioned; however, it holds potential for the future. In particular, the home-based nature of the program seems to contain possibilities for supporting individualization, autonomy, independence, and adherence in addition to supporting improvement in physical capability in older patients. Considering that some older patients are digital immigrants and may oppose new technology, prehabilitation can advantageously be implemented to support familiarity toward technologies and execution of exercise.

For the patients, having a good relationship with health professionals seemed important for their attitude toward the program, and continuity in meeting the same health professional more than once added to the positive experience of having a good relationship. According to research by Feo et al [43] and Bridges et al [44], a positive, trusting relationship supports dignity and helps patients to make informed decisions in addition to supporting the delivery of high-quality care. Therefore, it seemed beneficial to support continuity in the interactions between health professionals and patients when conducting exercise-based CTR.

Strengths and Limitations

This study has some limitations. First, the small number of participants resulted in reduced data saturation; however, rich data were gathered, which allowed for in-depth exploration of patients' and health professionals' experiences of participating in TeleTAVI [45]. Second, the interviews were conducted after 8 weeks of web-based training, meaning that it was impossible to gain insight into experiences of long-term adherence to training.

However, this study has several strengths. Despite the low number of participants, the group of patients reflected the population in the following areas: age, comorbidities, and education level [46]. Another strength is that all interviews were conducted in the patients' private homes to support a relaxed and comfortable atmosphere, to make the patients feel safe, and both interviewers were experienced in conducting semistructured interviews with patients with cardiac diseases. In addition, it is considered a strength that both patients and health professionals involved in the telerehabilitation were interviewed to ensure deep and more comprehensive understanding of the feasibility of the telerehabilitation program.

Finally, another strength of the study is that CBT and AV undertook in-depth reading and analysis to avoid misinterpretation of the interviews, and an ongoing discussion of the analysis and interpretation was conducted with the last author (BCB) to ensure rigor and credibility.

Conclusions

In conclusion, to the best of our knowledge, this was the first investigation of patients' and health professionals' experiences of exercise-based CTR for patients following TAVI surgery.

The feasibility of TeleTAVI in the current form may be questioned. The patients were reticent toward using the website, and they had neither the skills nor the curiosity to explore the tablet. They found the use of the activity tracker interesting, but full benefit was not achieved from it. The patients had low technical competencies, leading to feeling worried before the training, as they anticipated technical problems. The physiotherapists observed improvements in patients' physical fitness despite technical problems and that home training supported safety, individualization, and adherence. Individualization was valued by the patients, and they experienced physical improvements. Good relationships and continuity in contact with health professionals were extremely important for the patients, and they did not express any wish to be socially related to the other participants during the program.

However, TeleTAVI holds potential for the future. In particular, the home-based nature of the program seems to contain possibilities for supporting individualization, autonomy, independence, and adherence in addition to supporting improvement in physical capability in older patients. In addition, basing the relationship on continuity between health professionals and patients may be beneficial for patients. Considering that some older patients are digital immigrants and may oppose new technology, prehabilitation can advantageously be implemented to support familiarity toward technologies and execution of exercise.

Acknowledgments

The authors would like to thank the participating patients and health professionals for their valuable contributions.

Authors' Contributions

BCB and CBT were involved in conceptualization and methodology; BCB, J Andreasen, JJA, J Aarøe, and CBT contributed to protocol writing; BCB was involved in data collection and project administration; CBT, AV, and BCB contributed in data analyses, writing the paper, and preparation of the original draft; and BCB, J Andreasen, JJA, J Aarøe, AV, and CBT contributed in writing and reviewing the paper. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Informative website created for TeleTAVI.
[DOCX File , 1423 KB - formative_v6i6e35365_app1.docx]

References

1. Fishbein GA, Schoen FJ, Fishbein MC. Transcatheter aortic valve implantation: status and challenges. Cardiovasc Pathol 2014;23(2):65-70. [doi: 10.1016/j.carpath.2013.10.001] [Medline: 24183003]
2. Vahanian A, Beyersdorf F, Praz F, Milojevic M, Baldus S, Bauersachs J, ESC/EACTS Scientific Document Group. 2021 ESC/EACTS guidelines for the management of valvular heart disease. Eur J Cardiothorac Surg 2021 Oct 22;60(4):727-800. [doi: 10.1093/ejcts/ezab389] [Medline: 34453161]
3. Lytvyn L, Guyatt GH, Manja V, Siemieniuk RA, Zhang Y, Agoritsas T, et al. Patient values and preferences on transcatheter or surgical aortic valve replacement therapy for aortic stenosis: a systematic review. BMJ Open 2016 Sep 29;6(9):e014327 [FREE Full text] [doi: 10.1136/bmjopen-2016-014327] [Medline: 27687903]
4. von Kappelgaard L, Toftlund SA, Davidsen M, Gislason G. Dansk Hjerteregisters Årsberetning 2017. Statens Institut for Folkesundhed. 2018 Jun 27. URL: https://www.sdu.dk/da/sif/rapporter/2018/dansk_hjerteregisters_aarsberetning_2017 [accessed 2019-05-01]
5. Olsson K, Nilsson J, Hörnsten Å, Näslund U. Patients' self-reported function, symptoms and health-related quality of life before and 6 months after transcatheter aortic valve implantation and surgical aortic valve replacement. Eur J Cardiovasc Nurs 2017 Mar;16(3):213-221. [doi: 10.1177/1474515116650342] [Medline: 27169460]
6. Ribeiro GS, Melo RD, Deresz LF, Dal Lago P, Pontes MR, Karsten M. Cardiac rehabilitation programme after transcatheter aortic valve implantation versus surgical aortic valve replacement: systematic review and meta-analysis. Eur J Prev Cardiol 2017 May;24(7):688-697. [doi: 10.1177/2047487316686442] [Medline: 28071146]
7. Völler H, Salzwedel A, Nitardy A, Buhlert H, Treszl A, Wegscheider K. Effect of cardiac rehabilitation on functional and emotional status in patients after transcatheter aortic-valve implantation. Eur J Prev Cardiol 2015 May;22(5):568-574. [doi: 10.1177/2047487314526072] [Medline: 24577878]

8. Abraham LN, Sibilitz KL, Berg SK, Tang LH, Risom SS, Lindschou J, et al. Exercise-based cardiac rehabilitation for adults after heart valve surgery. Cochrane Database Syst Rev 2021 May 07;5(5):CD010876 [FREE Full text] [doi: 10.1002/14651858.CD010876.pub3] [Medline: 33962483]
9. Sathananthan J, Lauck S, Piazza N, Martucci G, Kim DH, Popma JJ, et al. Habitual physical activity in older adults undergoing TAVR: insights from the FRAILTY-AVR study. JACC Cardiovasc Interv 2019 Apr 22;12(8):781-789 [FREE Full text] [doi: 10.1016/j.jcin.2019.02.049] [Medline: 31000014]
10. Anderson L, Thompson DR, Oldridge N, Zwisler AD, Rees K, Martin N, et al. Exercise-based cardiac rehabilitation for coronary heart disease. Cochrane Database Syst Rev 2016 Jan 05;2016(1):CD001800 [FREE Full text] [doi: 10.1002/14651858.CD001800.pub3] [Medline: 26730878]
11. Hansen TB, Berg SK, Sibilitz KL, Søgaard R, Thygesen LC, Yazbeck AM, et al. Availability of, referral to and participation in exercise-based cardiac rehabilitation after heart valve surgery: results from the national CopenHeart survey. Eur J Prev Cardiol 2015 Jun;22(6):710-718. [doi: 10.1177/2047487314536364] [Medline: 24857890]
12. Parmanto B, Saptono A. Telerehabilitation: state-of-the-art from an informatics perspective. Int J Telerehabil 2009 Sep 4;1(1):73-84 [FREE Full text] [doi: 10.5195/ijt.2009.6015] [Medline: 25945164]
13. Brennan DM, Tindall L, Theodoros D, Brown J, Campbell M, Christiana D, American Telemedicine Association. A blueprint for telerehabilitation guidelines--October 2010. Telemed J E Health 2011 Oct;17(8):662-665. [doi: 10.1089/tmj.2011.0036] [Medline: 21790271]
14. Rawstorn JC, Gant N, Direito A, Beckmann C, Maddison R. Telehealth exercise-based cardiac rehabilitation: a systematic review and meta-analysis. Heart 2016 Aug 01;102(15):1183-1192. [doi: 10.1136/heartjnl-2015-308966] [Medline: 26936337]
15. Huang K, Liu W, He D, Huang B, Xiao D, Peng Y, et al. Telehealth interventions versus center-based cardiac rehabilitation of coronary artery disease: a systematic review and meta-analysis. Eur J Prev Cardiol 2015 Aug;22(8):959-971. [doi: 10.1177/2047487314561168] [Medline: 25488550]
16. Scherrenberg M, Falter M, Dendale P. Cost-effectiveness of cardiac telerehabilitation in coronary artery disease and heart failure patients: systematic review of randomized controlled trials. Eur Hear J 2020 Nov;1(1):20-29. [doi: 10.1093/ehjdh/ztaa005]
17. Frederix I, Vanhees L, Dendale P, Goetschalckx K. A review of telerehabilitation for cardiac patients. J Telemed Telecare 2015 Jan;21(1):45-53. [doi: 10.1177/1357633X14562732] [Medline: 25475219]
18. Xu L, Li F, Zhou C, Li J, Hong C, Tong Q. The effect of mobile applications for improving adherence in cardiac rehabilitation: a systematic review and meta-analysis. BMC Cardiovasc Disord 2019 Jul 12;19(1):166 [FREE Full text] [doi: 10.1186/s12872-019-1149-5] [Medline: 31299903]
19. Mehra S, Visser B, Cila N, van den Helder J, Engelbert RH, Weijs PJ, et al. Supporting older adults in exercising with a tablet: a usability study. JMIR Hum Factors 2019 Feb 01;6(1):e11598 [FREE Full text] [doi: 10.2196/11598] [Medline: 30707106]
20. Dinesen B, Spindler H. The use of telerehabilitation technologies for cardiac patients to improve rehabilitation activities and unify organizations: qualitative study. JMIR Rehabil Assist Technol 2018 Nov 19;5(2):e10758 [FREE Full text] [doi: 10.2196/10758] [Medline: 30455168]
21. Spindler H, Leerskov K, Joensson K, Nielsen G, Andreasen JJ, Dinesen B. Conventional rehabilitation therapy versus telerehabilitation in cardiac patients: a comparison of motivation, psychological distress, and quality of life. Int J Environ Res Public Health 2019 Feb 12;16(3):512 [FREE Full text] [doi: 10.3390/ijerph16030512] [Medline: 30759761]
22. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
23. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975 Nov;12(3):189-198. [doi: 10.1016/0022-3956(75)90026-6] [Medline: 1202204]
24. ATS Committee on Proficiency Standards for Clinical Pulmonary Function Laboratories. ATS statement: guidelines for the six-minute walk test. Am J Respir Crit Care Med 2002 Jul 01;166(1):111-117. [doi: 10.1164/ajrccm.166.1.at1102] [Medline: 12091180]
25. National Klinisk Retningslinje for Hjerterehabilitering. Sundhedsstyrelsen. 2015 Feb 20. URL: https://www.sst.dk/da/udgivelser/2015/nkr-hjerterehabilitering [accessed 2019-05-01]
26. Kvale S, Brinkmann S. InterViews: Learning the Craft of Qualitative Research Interviewing. Thousand Oaks, CA, USA: Sage Publications; May 01, 2014.
27. Vaismoradi M, Turunen H, Bondas T. Content analysis and thematic analysis: implications for conducting a qualitative descriptive study. Nurs Health Sci 2013 Sep;15(3):398-405. [doi: 10.1111/nhs.12048] [Medline: 23480423]
28. Zamawe FC. The implication of using NVivo software in qualitative data analysis: evidence-based reflections. Malawi Med J 2015 Mar;27(1):13-15 [FREE Full text] [doi: 10.4314/mmj.v27i1.4] [Medline: 26137192]
29. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004 Feb;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]

30. Andreasen J, Sørensen EE, Gobbens RJ, Lund H, Aadahl M. Danish version of the Tilburg Frailty Indicator--translation, cross-cultural adaption and validity pretest by cognitive interviewing. Arch Gerontol Geriatr 2014;59(1):32-38. [doi: 10.1016/j.archger.2014.02.007] [Medline: 24680192]
31. Gobbens RJ, Boersma P, Uchmanowicz I, Santiago LM. The tilburg frailty indicator (TFI): new evidence for its validity. Clin Interv Aging 2020 Feb 21;15:265-274 [FREE Full text] [doi: 10.2147/CIA.S243233] [Medline: 32110005]
32. Fischer SH, David D, Crotty BH, Dierks M, Safran C. Acceptance and use of health information technology by community-dwelling elders. Int J Med Inform 2014 Sep;83(9):624-635 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.06.005] [Medline: 24996581]
33. Terp M. Exploring the possibilities of smartphone-based young adult schizophrenia care: a participatory design study. Aalborg Universitetsforlag. 2019. URL: https://vbn.aau.dk/ws/portalfiles/portal/306525393/PHD_Malene_Terp_E_pdf.pdf [accessed 2021-08-12]
34. Hargittai E, Piper AM, Morris MR. From Internet access to Internet skills: digital inequality among older adults. Univ Access Inf Soc 2018 May 3;18(4):881-890. [doi: 10.1007/s10209-018-0617-5]
35. van Leeuwen KM, van Loon MS, van Nes FA, Bosmans JE, de Vet HC, Ket JC, et al. What does quality of life mean to older adults? A thematic synthesis. PLoS One 2019 Mar 8;14(3):e0213263 [FREE Full text] [doi: 10.1371/journal.pone.0213263] [Medline: 30849098]
36. Carstensen G, Rosberg B, Mc Kee KJ, Åberg AC. Before evening falls: perspectives of a good old age and healthy ageing among oldest-old Swedish men. Arch Gerontol Geriatr 2019;82:35-44 [FREE Full text] [doi: 10.1016/j.archger.2019.01.002] [Medline: 30711841]
37. Walsh DM, Moran K, Cornelissen V, Buys R, Claes J, Zampognaro P, et al. The development and codesign of the PATHway intervention: a theory-driven eHealth platform for the self-management of cardiovascular disease. Transl Behav Med 2019 Jan 01;9(1):76-98. [doi: 10.1093/tbm/iby017] [Medline: 29554380]
38. French DP, Olander EK, Chisholm A, Mc Sharry J. Which behaviour change techniques are most effective at increasing older adults' self-efficacy and physical activity behaviour? A systematic review. Ann Behav Med 2014 Oct;48(2):225-234. [doi: 10.1007/s12160-014-9593-z] [Medline: 24648017]
39. Frank AW. The Wounded Storyteller: Body, Illness, and Ethics. 2nd Edition. Chicago, IL, USA: The University of Chicago Press; May 05, 2013.
40. Eichler S, Salzwedel A, Reibis R, Nothroff J, Harnath A, Schikora M, et al. Multicomponent cardiac rehabilitation in patients after transcatheter aortic valve implantation: predictors of functional and psychocognitive recovery. Eur J Prev Cardiol 2017 Feb;24(3):257-264. [doi: 10.1177/2047487316679527] [Medline: 27852810]
41. Pighi M, Piazza N, Martucci G, Lachapelle K, Perrault LP, Asgar AW, et al. Sex-specific determinants of outcomes after transcatheter aortic valve replacement. Circ Cardiovasc Qual Outcomes 2019 Mar;12(3):e005363. [doi: 10.1161/CIRCOUTCOMES.118.005363] [Medline: 30879326]
42. Pressler A, Förschner L, Hummel J, Haller B, Christle JW, Halle M. Long-term effect of exercise training in patients after transcatheter aortic valve implantation: follow-up of the SPORT:TAVI randomised pilot study. Eur J Prev Cardiol 2018 May;25(8):794-801. [doi: 10.1177/2047487318765233] [Medline: 29553289]
43. Feo R, Conroy T, Marshall RJ, Rasmussen P, Wiechula R, Kitson AL. Using holistic interpretive synthesis to create practice-relevant guidance for person-centred fundamental care delivered by nurses. Nurs Inq 2017 Apr;24(2):e12152. [doi: 10.1111/nin.12152] [Medline: 27549018]
44. Bridges J, Nicholson C, Maben J, Pope C, Flatley M, Wilkinson C, et al. Capacity for care: meta-ethnography of acute care nurses' experiences of the nurse-patient relationship. J Adv Nurs 2013 Apr;69(4):760-772 [FREE Full text] [doi: 10.1111/jan.12050] [Medline: 23163719]
45. Malterud K, Siersma VD, Guassora AD. Sample size in qualitative interview studies: guided by information power. Qual Health Res 2016 Nov;26(13):1753-1760. [doi: 10.1177/1049732315617444] [Medline: 26613970]
46. Borregaard B, Sørensen J, Ekholm O, Møller JE, Riber LP, Thrysoee L, et al. Sociodemographic, clinical and patient-reported outcomes and readmission after heart valve surgery. J Heart Valve Dis 2018 Jan;27(1):78-86. [Medline: 30560603]

Abbreviations

AS: aortic valve stenosis
CR: cardiac rehabilitation
CTR: cardiac telerehabilitation
ICT: information and communications technology
TAVI: transcatheter aortic valve implantation

Edited by T Leung; submitted 14.02.22; peer-reviewed by N Maglaveras, N Tenbult; comments to author 29.03.22; revised version received 29.04.22; accepted 16.05.22; published 20.06.22.

Please cite as:
Thorup CB, Villadsen A, Andreasen JJ, Aarøe J, Andreasen J, Brocki BC
Perspectives on Participation in a Feasibility Study on Exercise-Based Cardiac Telerehabilitation After Transcatheter Aortic Valve Implantation: Qualitative Interview Study Among Patients and Health Professionals
JMIR Form Res 2022;6(6):e35365
URL: https://formative.jmir.org/2022/6/e35365
doi:10.2196/35365
PMID:35723925

Original Paper

Exploring Whether Addictions Counselors Recommend That Their Patients Use Websites, Smartphone Apps, or Other Digital Health Tools to Help Them in Their Recovery: Web-Based Survey

Tyler B Wray[1], PhD

Department of Behavioral and Social Sciences, Brown University School of Public Health, Providence, RI, United States

Corresponding Author:
Tyler B Wray, PhD
Department of Behavioral and Social Sciences
Brown University School of Public Health
121 S. Main Street
Box G-S121-5
Providence, RI, 02903
United States
Phone: 1 14018636659
Email: tyler_wray@brown.edu

Abstract

Background: Hundreds of smartphone apps or websites claiming to help those with addictions are available, but few have been tested for efficacy in changing clinically relevant addictions outcomes. Although most of these products are designed for self-facilitation by users struggling with addictions, counselors and other addictions treatment providers will likely play a critical role in facilitating adoption by integrating their use into counseling or recommending them to their patients. Yet, few studies have explored the practices of addictions counselors in using or recommending addictions-focused digital health tools in their work.

Objective: The aim of this study was to understand whether addiction counselors are recommending that their patients use addictions-focused apps to help them in their recovery, and the factors that affect their desire to do so.

Methods: Licensed addiction counselors practicing in the United States (N=112) were recruited from professional and scientific organizations of alcohol or drug counselors to complete a web-based survey.

Results: In total, 74% (83/112) of counselors had recommended that their patients use a website or smartphone app to assist them in recovery, and those that had done so reported recommending an app with an average of 54% of their patients. The most commonly recommended app or website was SMARTRecovery.org (9%), I am Sober (8%), In the Rooms (7%), Insight Timer (4%), Calm (4%), Sober Tool (4%), Recovery Box (3%), and Sober Grid (3%). The most important reason that counselors recommended the websites or apps was that colleagues or patients told them they found it helpful (55%), followed by their workplaces recommending it (20%) and professional organizations recommending it (10%). Counselors' intentions to recommend a hypothetical app were strongest for apps that had been tested in rigorous, scientific studies that showed they helped users stay sober or reduce their substance use; 94% (105/112) reported that they would "definitely" or "probably" use such an app.

Conclusions: Most addictions counselors surveyed are already recommending that their patients use apps or websites to help them in their recovery, despite the paucity of available products that have evidence supporting their efficacy for addictions outcomes. One way that product developers could increase adoption among addictions treatment providers is to make efficacy testing a priority and to disseminate results through professional organizations and clinics.

(JMIR Form Res 2022;6(6):e37008) doi:10.2196/37008

KEYWORDS

addiction; alcohol; drug use; substance use; adoption; smartphone; mobile health; mHealth; marketing; dissemination; counselor; health care professional; digital health; eHealth

Introduction

The market of health-related digital tools, such as websites and smartphone apps, is continuing to grow exponentially, with more than 90,000 health apps added to major app stores in 2020 [1]. Total funding for digital health ventures also surpassed US $21.3 billion in the first three-quarters of 2021, up from US $1.1 billion just a decade ago [2]. Mental health–focused products make up the largest share of apps available to help manage health conditions [1], and several mental wellness apps (eg, Calm and Headspace) were represented among the most downloaded apps in 2020 [3].

Although hundreds of addictions-focused apps are currently available online and in app stores, only a handful of these products have been formally tested for their efficacy or effectiveness in changing addiction-relevant outcomes such as alcohol or drug use and problems or substance use disorder symptoms or severity [4]. Similarly, only a few addictions-focused apps have sought premarket clearances or approvals that require them to submit data on their products' safety or efficacy from controlled studies [1]. This means that evidence supporting clinical benefit is only available to the public for a very small number of addictions-focused apps. However, a number of such products have been developed and have shown promising effects on a variety of addictions outcomes [5]. A handful of products have also recently sought Food and Drug Administration approval or clearance for their addictions-focused products to begin marketing their benefits for those in recovery (eg, reSET [6]). Although no treatment guidelines or major professional organizations have yet explicitly suggested that clinicians use or recommend any such products, it is very likely that one or more of these products will prove beneficial enough to achieve this degree of support at some point in the near future.

Most of the addiction-focused websites and smartphone apps that have been developed thus far have been primarily designed for self-facilitation, meaning that patients are primarily intended to use them on their own, privately [7]. For this reason, addiction-focused products, such as other mental health–focused apps, often pursue at least some patient-focused marketing strategies as a part of their overall marketing plans. These strategies, such as direct-to-consumer advertising, would likely be more successful if they highlighted any benefits their products have shown on addictions outcomes in controlled research. That is, conducting research evaluating the efficacy of these apps on clinically relevant outcomes (such as alcohol or drug use and problems or substance use disorder symptoms) may be important for increasing confidence and interest in the product among consumers. However, the uptake of these products among patients may also be higher when treatment providers recommend that they use them [8], leading many product developers in other spaces to pursue marketing their consumer-focused digital health products through medical and mental health providers. Some addiction-focused products have also pursued regulatory clearance or approval under a prescription-based model (eg, reSET [6]), meaning that patient uptake is not possible at all without a provider's recommendation. Other products, such as virtual reality experiences [9], may also be explicitly intended to be used collaboratively by counselors and patients during treatment sessions, which is another scenario that requires counselor adoption in order for the product to be used by patients. Given these approaches, understanding the practices of addictions providers in recommending websites or apps to their patients and factors affecting their likelihood of recommending them is critical for successfully disseminating addiction-focused digital health products. Few such studies have been conducted to date.

In this study, I explored whether licensed, practicing addictions counselors are currently recommending that their patients use websites or smartphone apps to help them in their recovery, and examined some factors that could affect their desire to do so. I also examined marketing or dissemination content and venues that could be most likely to reach and encourage addiction counselors to recommend websites or apps to their patients.

Methods

Participants

Participants (N=112) were recruited from email groups, listserv postings, and advertisements in newsletters maintained by professional and scientific organizations of drug or alcohol counselors, as well as social media posts and advertisements, to participate in a web-based survey from February to July 2021. Eligible participants (1) were at least 18 years old, (2) were able to read fluently in English, (3) were licensed to provide counseling in the United States, and (4) provided alcohol or drug counseling for at least 15% of their typical work week. Moreover, another manuscript that is currently in review reported on other data from this survey (TB Wray, unpublished data, September 2021).

Procedures

The participants first completed a web-based screening survey to determine eligibility. If the respondents were eligible and interested, they were asked to provide informed consent and contact information before being redirected to the full survey. The main survey took participants an average of about 25 minutes to complete. Those participants who completed the full survey were compensated with a US $20 gift card, sent via email.

Ethics Approval

All procedures were reviewed by the Brown University IRB and were determined to be exempt from ongoing review and approval (protocol # 2101002892), because the study procedures only involved a single survey that collected data on a topic that, if these data were inadvertently disclosed, would not reasonably place the subjects at risk of criminal or civil liability or damage their financial standing, employability, educational advancement, or reputation.

Measures

Since few validated measures have been developed to assess counselors' perceptions about digital health apps, most items were created for this study. The question about what websites or app counselors had recommended allowed respondents to enter text freely. To evaluate whether each product had been

the focus of published, peer-reviewed research, I searched several databases (eg, Google Scholar, PubMed, and Food and Drug Administration's Premarket Approval 510k databases, if applicable) for each by their listed brand name, as well as websites that the product or its developer maintained for relevant research reports. Questions about counselors' *perceptions of their patients' use* of addictions-focused apps were assessed on a 1 (*not at all*) to 5 (*a lot*) scale. Questions asking counselors to estimate the *percentage of their patients they recommended appsto* were rated on a slider scale from 1% to 100%. Items asking the counselors to rate their intentions to recommend a hypothetical app under various conditions (eg, if rigorous, scientific studies had shown it was effective) were rated on a 1 (*not at all*) to 5 (*definitely yes*) scale. The question asking about the counselors' most important reservations about recommending digital health products to their patients displayed 9 potential reservations (e., "I don't know any that help with that, they would probably cost too much") plus an "other" option with free text entry. The participants were instructed to select those reservations that were concerns for them, and then rank them from most to least important. This item was displayed only to participants who had not ever recommended an app or website and expressed at least some reservation about doing so in the future (ie, responded lower than "definitely yes" on the item asking about their intentions to recommend any app or website to their patients in the future; N=26).

Data Analysis Plan

Data from all complete responses for the items reported were included in these analyses. Descriptive statistics were calculated for all demographic and professional characteristics. Basic summary statistics were used (percentages, medians, and interquartile range for Likert scale data) for all focal study items. All analyses were conducted in Stata 16 (StataCorp).

Results

Participant demographic characteristics are presented in Table 1, and professional characteristics are presented in Table 2. The respondents (N=112) were largely master-level clinicians (n=72, 64.3%). Nearly half (n=52, 46.4%) were licensed substance abuse counselors, with many social workers (n=29, 25.9%) and mental health counselors (n=20, 17.9%) as well. Most participants reported working in an outpatient addiction treatment center, in private practice, outpatient mental health centers, or criminal justice settings. According to national data on the substance use disorder treatment workforce [10,11], a higher percentage of participants in this study were counseling staff (99% vs 42%), had a graduate education (75% vs 57%), and earned a higher annual salary (US $71,937 vs US $48,520) compared to the average substance use disorder treatment professional in the United States.

Table 1. Demographic characteristics of the participants (N=112).

Characteristics	Values
Age (years), mean (SD; range)	45.6 (12.3; 25-76)
Sex (female)	86 (76.8)
Race or ethnicity, n (%)	
White	93 (83.0)
Black or African American	13 (11.6)
Asian	2 (1.8)
American Indian or Alaska Native	0 (0)
Multiracial	4 (3.6)
Hispanic or Latino	5 (4.5)
Sexual orientation, n (%)	
Heterosexual or straight	96 (85.7)
Gay or lesbian	7 (6.3)
Bisexual	7 (6.3)
Other or chose not to respond	2 (1.8)
US region of primary residence, n (%)	
Northeast	55 (49.1)
South	36 (32.1)
Midwest	13 (11.6)
West	8 (7.1)

Table 2. Professional characteristics of the participants (N=112).

Characteristics	Values
Education, n (%)	
Some college	1 (0.9)
Bachelor's degree	22 (19.6)
Some grad school	5 (4.5)
Master's degree	72 (64.3)
Some doctoral work	3 (3.0)
Doctorate degree (PhD, PsyD, DNP)	8 (7.1)
Medical doctorate (MD, DO)	1 (0.9)
Annual income (US $), mean (SD)	71,937 (23,618)
Employed full-time, n (%)	104 (92.9)
Professional discipline, n (%)	
Psychologist	6 (5.4)
Social worker	29 (25.9)
Substance abuse counselor	52 (46.4)
Health educator	2 (1.8)
Mental health counselor	20 (17.9)
School counselor	1 (0.9)
Physician (other than psychiatry)	1 (0.9)
Other	1 (0.9)
Years of training in addictions counseling, mean (SD)	3.6 (2.1)
Years of experience in addictions counseling, mean (SD)	10.4 (6.4)
Number of sessions with each addictions client, mean (SD)	14 (8.8)
Primary work setting, n (%)	
Community hospital	4 (3.6)
Academic hospital	6 (5.4)
Veterans' affairs or military hospital	5 (4.5)
Outpatient mental health center	16 (14.3)
Inpatient mental health center	5 (4.5)
Outpatient addiction treatment center	21 (18.8)
Inpatient addiction treatment center	4 (3.6)
Prison or detention center	16 (14.3)
Community medical clinic or health center	7 (6.3)
Private medical clinic or health center	1 (0.9)
Government or social service center	2 (1.8)
K-12 school	2 (1.8)
Private practice	19 (16.0)
Other	2 (1.8)
Primary counseling approach, n (%)	
12-step facilitation	5 (4.5)
Cognitive behavioral therapy	15 (13.4)
Motivational interviewing	51 (45.5)
Acceptance and commitment therapy	7 (6.3)

Characteristics	Values
Mindfulness-based relapse prevention	3 (2.7)
Dialectical behavior therapy	2 (1.8)
Narrative therapy	3 (2.7)
Harm reduction	2 (1.8)
Eclectic	8 (7.1)
Other	16 (14.3)

In total, 74% (83/112) of the respondents had recommended that their patients with addiction use a website or smartphone app to assist them in recovery. Among those who had ever recommended a smartphone app or website, participants reported recommending 122 different apps or websites (Table 3). These participants also reported recommending one of these apps to an average of 54% of their patients. Unsurprisingly, the most recommended category of website or app was general recovery support apps, which provide a wide variety of features and content intended to facilitate recovery, including forums or groups to provide social support; meeting or service locators; daily meditations; and behavior, mood, and thought tracking, among others. One such platform, SMARTRecovery.org (9%, 11/122) was the most recommended application overall. Meditation and mindfulness apps were the second most common type of website or app that counselors recommended to their patients, with Calm (4%, 5/122) and Insight Timer (4%, 5/122) among the most popular of these. Sobriety trackers were the next most recommended category of website or app, and I Am Sober (7%, 9/122) was among the most recommended overall. In the Rooms, which is a 12-step facilitation app that links users with Alcoholics Anonymous and Narcotics Anonymous meetings nearby was also among the most common recommendations overall (7%, 9/122). Finally, some counselors also reported recommending that their patients use websites or apps that provide videos, podcasts, or written text that deliver recovery-related education and inspiration. Among the most popular of these were the Hazelden suite of smartphone apps, such as Twenty-Four Hours a Day, which provides written affirmational or inspirational content to users. Only 1 of the top 10 most frequently recommended websites or apps, Calm, has been the focus of published, peer-reviewed efficacy or effectiveness research exploring the effects of that specific product on outcomes that were potentially related to addictions. None of the apps that the counselors recommended have yet been the focus of efficacy research specifically on addiction outcomes, such as alcohol or drug use and problems, or substance use disorder symptoms.

The most frequently reported reason that counselors recommended the websites or apps to their patients was that colleagues or patients had told them they found it helpful (55%, 45/82), followed by their places of work recommending it (20%, 16/82) and professional organizations or groups recommending it (10%, 8/82). Only 9% (7/82) of the counselors reported recommending a particular website or app to their patients because rigorous, scientific studies had shown that it was helpful. However, all counselors reported the strongest intentions to recommend that their patients use a website or app to help them in recovery if rigorous, scientific studies (eg, randomized controlled trials) had been conducted on the efficacy of that website or app and had shown that using it helped patients stay sober or reduce their substance use; 94% (105/112) reported that they would “definitely” or “probably” recommend a product that was supported by this degree of evidence. The counselors also had high intentions to recommend websites or apps if others they knew (eg, colleagues or patients) found it helpful; 92% (103/112) reported that they would “definitely” or “probably” recommend such a product. Finally, only 38% (43/112) of the counselors reported that they would “definitely” or “probably” recommend an app or website to patients if its developers claimed it was helpful for users in recovery.

Of counselors who had recommended at least one app (82/112, 74%), the most common way they heard about the apps they recommended was web-based search, followed by other counselors or clinicians, patients, and professional organization conferences or conventions. By contrast, most participants reported that they would *like* to hear about addiction-focused apps through other counselors or clinicians and patients, followed by hearing about them through conferences or publications (eg, newsletters) maintained by their professional organizations, or through seminars at work (Table 4).

Of the 30 counselors who had not ever recommended a website or app, 79% (n=24) said they would either “probably” or “definitely” recommend that their patients use a website or smartphone app to help their patients stay sober or reduce their alcohol or drug use. However, the most frequently chosen reservation among those who had not recommended an app or were uncertain about doing so was that they did not know of any app that had been shown scientifically to help with addictions (11.6%). Other frequently identified and highly ranked reservations were concerns that their patients would not actually use them, not being generally aware of any websites or apps, and believing that their patients do not have the resources to use a computer or a smartphone (Table 5).

Table 3. Apps and websites that addiction counselors most recommended to their patients in recovery and their core functions (only categories with >1 recommendation are represented).

App or website category	Description[a]	Percent[b]	Most common sites or apps[b] (%)
General recovery support	Provides several features, such as forums, meeting or service locators, daily meditations, sobriety counters, and calculators, intended to provide general recovery support.	21.3	• SMARTRecovery.org (9) • RecoveryBox (2) • Sober Grid (2)
Meditation or mindfulness	Provides content to help users develop mindfulness skills, such as guided meditations, education, and badges or rewards.	19.7	• Calm (4) • Insight Timer (4) • Headspace (2)
Sobriety trackers	Primarily helps users track their length of sobriety, including various calculators of benefit (eg, money saved)	18.0	• I Am Sober (7) • Sober Tool (4) • Nomo (2)
12-step facilitation	Helps connect users to AA[c] or NA[d] meetings or hosts meetings on their platforms and other AA- or NA-associated content.	14.8	• In the Rooms (7) • Meeting Guide (1) • 12 Step Companion (1)
Inspiration, podcasts, or lectures	Provides video, audio, or text on recovery-related topics primarily to educate or motivate (inspire) users.	7.4	• Hazelden Recovery Apps (4) • SoberCast (2) • YouTube (2)
Social media	General social media sites or apps; counselors recommended specific recovery communities and groups.	7.4	• Facebook (2) • Reddit (2)
General mental health	Provides content intended to support general mental health, such as mood tracking and thought diaries.	2.5	• CBT Thought Diary (1) • Mood Meter (1) • Woebot (1)
Smoking cessation	Provides support for smoking cessation, such as quit planning, nicotine replacement therapy, and social support.	2.5	• smokefree.org (1) • NY Quits (1) • QuitlineNC (1)
Unknown or other	App or website did not fit into other categories, was not found, or its purpose was unclear.	4.9	• Strengths Finder (1) • Weconnect (1)

[a]Apps or websites were grouped according to their general purposes. Many apps or websites across categories contain similar sets of features (eg, social support and daily meditations).

[b]All percentages represent the percent of all identified apps or websites that the participants reported recommending.

[c]AA: Alcoholics Anonymous.

[d]NA: Narcotics Anonymous.

Table 4. The platform from which counselors first heard of the addiction-focused apps they recommended, and where they would like to hear about them[a].

Platform or person	*Did* hear (N=82), frequency (%)	*Would like* to hear (N=112), frequency (%)
Web-based search	39 (34.8)	27 (24.1)
Other counselors or clinicians	36 (32.1)	75 (67.0)
Other patients or patients	23 (20.5)	49 (43.8)
Professional organization conference or convention	15 (13.4)	38 (33.9)
Professional organization newsletter or bulletin	14 (12.5)	44 (39.3)
App store (Apple App Store or Google Play)	12 (10.7)	11 (9.8)
Workshop or seminar at work	11 (9.8)	34 (30.4)
Other	7 (6.3)	0 (0)
Scholarly journal article	3 (2.7)	20 (17.9)
TV advertisement	1 (0.9)	6 (5.4)
Internet advertisement	0 (0)	10 (8.9)
Company press release	0 (0)	0 (0)

[a]Participants could select multiple ways in which they heard about the apps or websites they recommended. These data represent the number of times each method was identified and the percentage of all responses in which that method was selected.

Table 5. Addiction counselors' most frequently chosen reservations about recommending that their patients use websites or apps to help them with their recovery[a] (N=26).

Comments	Frequency[b] (%)	Rank, median (IQR)
I don't know of any websites or apps that help with addictions.	9 (8.0)	2 (1;3)
I don't know of any that have been shown scientifically to help with addictions.	13 (11.6)	2 (1;2)
I don't think websites or apps can help people reduce their alcohol or drug use.	5 (4.5)	4 (2;7)
They would probably cost too much.	11 (9.82)	2 (2;3)
I don't think my patients would want to use them.	7 (6.3)	3 (2;4)
I don't think my patients would actually use it even if they wanted to.	10 (10.7)	2 (1;3)
I don't think they have the resources (a computer or smartphone) to use them.	9 (8.0)	1 (1;2)
I'm concerned it may do more harm than good	7 (6.3)	3 (2;5)
I'm concerned that, if I did, they may stop coming to counseling.	4 (3.6)	6.5 (3;8.5)
Other	3 (2.7)	3 (3;4)

[a]This item was only rated by respondents who expressed never having recommended an app to a patient in recovery *and* expressed some reservation about doing so in the future (N=26).

[b]The respondents could select multiple reservations. Frequencies reflects the number of times each reason was chosen and the percentage of all selections.

Discussion

Principal Findings

The results of this study showed that 3 (75%) of every 4 licensed practicing addictions counselors surveyed had recommended that a client use a website or smartphone app to assist them in their recovery. The results also showed that counselors who had recommended an app did so with about half of their patients, suggesting that many counselors are already actively recommending addictions-focused or mental health–focused digital health products to most of their patients. However, only 1 of the apps that the counselors listed, Calm—a mindfulness and meditation app primarily marketed for improving sleep and reducing stress—has been the focus of peer-reviewed, published efficacy research, which included outcomes that are potentially relevant to addictions to date. Randomized controlled trials have shown that using the Calm app improves sleep outcomes among adults with sleep disturbance [12]. Others have shown that it could reduce depression and anxiety among patients with cancer [13,14], and that it could reduce stress among college students [15]. Although none of these research projects focuses on addictions outcomes specifically, sleep disturbances and mental health challenges frequently co-occur with substance use disorders [16,17], and counselors may be recommending the Calm app to patients to help them with those conditions. None of the apps that the counselors reported recommending to their patients had been the focus of published, peer-reviewed efficacy research focusing on outcomes directly related to addictions. However, one of the most frequently recommended applications,

SMARTRecovery.org (Self-Management and Recovery Training), is based on a face-to-face, group-based approach to addictions treatment, and this nondigital treatment has shown some promise in preliminary studies [18]. SMARTRecovery.org connects users with online group meetings and other web-based tools (eg, worksheets and exercises), but the efficacy of these web-based tools on addictions outcomes has not yet been evaluated [19]. Overall, these findings suggest that the desire of counselors is high enough to provide patients who have addiction with tools that can help them in their recovery; counselors may be recommending apps to them despite the limited availability of evidence about whether any of these apps are helpful in changing addictions-focused outcomes, such as alcohol or drug use and problems, or symptoms of substance use disorder. This is further supported by results suggesting that most counselors found the apps they recommended via online search, suggesting that many are actively searching for solutions for their patients on their own. Although the apps that the counselors recommended may be of some benefit for some patients and are unlikely to impair their recovery, doing so is not without risk; patients and counselors may spend money and time on solutions that are not helpful, and doing so could reduce trust in treatment generally. However, these findings also affirm that counselors could play an important role in encouraging the uptake of evidence-based, addictions-focused digital health tools generally.

The findings of this study also suggest that providing a strong evidence base for addictions-focused digital health products could increase the number of counselors who recommend these products to their patients. Across all counselors, intentions to recommend an app or website to their patients were highest if rigorous, scientific studies had shown it was helpful in reducing addictions outcomes. Many developers often consider efficacy testing to be a priority for helping achieve their regulatory goals; however, these results suggest that efficacy testing may also benefit the marketing and dissemination of addictions-focused products by increasing the number of counselors who recommend it to their patients. The results also suggested that counselors' intentions to recommend were also high for a hypothetical app that their colleagues or patients found helpful, which could suggest the importance of including endorsements and testimonials from trusted sources in marketing materials. This conclusion is further supported by the counselors' preferences that they hear about addictions-focused apps or websites through colleagues or patients. However, counselors also indicated that they would prefer to hear more about addictions-focused apps or websites from the professional associations they are involved in or their workplaces; this suggested that marketing efforts focused on both professional associations and large care providers could play an important role in successfully marketing addictions-focused digital health products. It could also suggest that implementation approaches that involve steps such as identifying and training local opinion leaders, early adopters, and "champions" of the intervention among counselors in specific professional organizations or treatment centers may be a particularly potent strategy for encouraging the uptake of any future evidence-based addictions-focused website or app [20].

Limitations

Several limitations are important to note. First, this sample represents a relatively small subset of addictions care providers; therefore, findings could be different in a larger sample or a sample composed of participants from a different variety of professional backgrounds. Second, the study materials (eg, recruitment advertisements and consent documents) disclosed that this study was about the role of technology in addictions counseling (as the institutional review board required it), so this may have attracted participants who are generally interested in using technology in addictions care. Therefore, the study may have overestimated the number of counselors who had or were willing to recommend apps to their patients. Third, the participants in the study were primarily of a White, non-Hispanic racial or ethnic background and were master-level clinicians. Therefore, the results reported here may not be generalizable to addictions counselors with other racial or ethnic backgrounds or education levels. Lastly, all items used in this study were created specifically for this study, so their reliability or validity has not yet been established.

Conclusions

In conclusion, the results of this study showed that a large majority of licensed, practicing addictions counselors had recommended apps or websites to their patients to help them in their recovery, and that those who had, did so with most of their patients. However, none of the apps that the counselors recommended to their patients had been supported by published, peer-reviewed, scientific research about their impact on addictions outcomes, likely because few such products are broadly available or marketed currently. However, the findings also showed that intentions to recommend a hypothetical addictions-focused digital health product were highest when rigorous research was available showing that the product helped reduce clinically relevant addiction outcomes, which in turn suggests that conducting this research could encourage counselors to recommend these products.

Acknowledgments

This manuscript was supported by L30AA023336 to TW from the National Institute on Alcohol Abuse and Alcoholism.

Conflicts of Interest

None declared.

References

1. Aitken M, Nass D. Digital health trends 2021: Innovation, evidence, regulation, and adoption. Slideshare. URL: https://www.slideshare.net/RicardoCaabate/digital-health-trends-2021-iqvia-global [accessed 2022-06-08]
2. Hawks C, Krasniansky A. Q3 2021 digital health funding: To $20B and beyond!. Rock Health. 2021 Oct 04. URL: https://rockhealth.com/insights/q3-2021-digital-health-funding-to-20b-and-beyond [accessed 2021-11-15]
3. Blacker A. Apptopia. 2021 Jan 07. URL: https://blog.apptopia.com/worldwide-us-download-leaders-2020 [accessed 2021-11-15]
4. Tofighi B, Chemi C, Ruiz-Valcarcel J, Hein P, Hu L. Smartphone Apps Targeting Alcohol and Illicit Substance Use: Systematic Search in in Commercial App Stores and Critical Content Analysis. JMIR Mhealth Uhealth 2019 Apr 22;7(4):e11831 [FREE Full text] [doi: 10.2196/11831] [Medline: 31008713]
5. Staiger PK, O'Donnell R, Liknaitzky P, Bush R, Milward J. Mobile Apps to Reduce Tobacco, Alcohol, and Illicit Drug Use: Systematic Review of the First Decade. J Med Internet Res 2020 Nov 24;22(11):e17156 [FREE Full text] [doi: 10.2196/17156] [Medline: 33231555]
6. Velez FF, Colman S, Kauffman L, Ruetsch C, Anastassopoulos K. Real-world reduction in healthcare resource utilization following treatment of opioid use disorder with reSET-O, a novel prescription digital therapeutic. Expert Rev Pharmacoecon Outcomes Res 2021 Feb 04;21(1):69-76. [doi: 10.1080/14737167.2021.1840357] [Medline: 33146558]
7. Goldfine C, Lai JT, Lucey E, Newcomb M, Carreiro S. Wearable and Wireless mHealth Technologies for Substance Use Disorder. Curr Addict Rep 2020 Sep 01;7(3):291-300 [FREE Full text] [doi: 10.1007/s40429-020-00318-8] [Medline: 33738178]
8. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
9. Powers MB, Emmelkamp PM. Virtual reality exposure therapy for anxiety disorders: A meta-analysis. J Anxiety Disord 2008 Apr;22(3):561-569. [doi: 10.1016/j.janxdis.2007.04.006] [Medline: 17544252]
10. Bouchery E, Dey J. DC: US Department of Health, Human Services OotASfP, Evaluation. Substance use disorder workforce issue brief. 2018. URL: https://aspe.hhs.gov/reports/substance-use-disorder-workforce-issue-brief-0 [accessed 2022-06-08]
11. Substance Abuse, Behavioral Disorder, and Mental Health Counselors. US Bureau of Labor Statistics. 2022. URL: https://www.bls.gov/ooh/community-and-social-service/substance-abuse-behavioral-disorder-and-mental-health-counselors.htm [accessed 2022-04-29]
12. Huberty JL, Green J, Puzia ME, Larkey L, Laird B, Vranceanu A, et al. Testing a mindfulness meditation mobile app for the treatment of sleep-related symptoms in adults with sleep disturbance: A randomized controlled trial. PLoS One 2021 Jan 7;16(1):e0244717 [FREE Full text] [doi: 10.1371/journal.pone.0244717] [Medline: 33411779]
13. Huberty J, Eckert R, Larkey L, Kurka J, Rodríguez De Jesús SA, Yoo W, et al. Smartphone-Based Meditation for Myeloproliferative Neoplasm Patients: Feasibility Study to Inform Future Trials. JMIR Form Res 2019 Apr 29;3(2):e12662 [FREE Full text] [doi: 10.2196/12662] [Medline: 31033443]
14. Puzia ME, Huberty J, Eckert R, Larkey L, Mesa R. Associations Between Global Mental Health and Response to an App-Based Meditation Intervention in Myeloproliferative Neoplasm Patients. Integr Cancer Ther 2020 Jun 20;19:1534735420927780 [FREE Full text] [doi: 10.1177/1534735420927780] [Medline: 32564631]
15. Huberty J, Green J, Glissmann C, Larkey L, Puzia M, Lee C. Efficacy of the Mindfulness Meditation Mobile App "Calm" to Reduce Stress Among College Students: Randomized Controlled Trial. JMIR Mhealth Uhealth 2019 Jun 25;7(6):e14273 [FREE Full text] [doi: 10.2196/14273] [Medline: 31237569]
16. Conroy DA, Arnedt JT. Sleep and substance use disorders: an update. Curr Psychiatry Rep 2014 Oct 19;16(10):487. [doi: 10.1007/s11920-014-0487-3] [Medline: 25135784]
17. Regier DA. Comorbidity of Mental Disorders With Alcohol and Other Drug Abuse. JAMA 1990 Nov 21;264(19):2511. [doi: 10.1001/jama.1990.03450190043026]
18. Zemore SE, Lui C, Mericle A, Hemberg J, Kaskutas LA. A longitudinal study of the comparative efficacy of Women for Sobriety, LifeRing, SMART Recovery, and 12-step groups for those with AUD. J Subst Abuse Treat 2018 May;88:18-26 [FREE Full text] [doi: 10.1016/j.jsat.2018.02.004] [Medline: 29606223]
19. Beck AK, Forbes E, Baker AL, Kelly PJ, Deane FP, Shakeshaft A, et al. Systematic review of SMART Recovery: Outcomes, process variables, and implications for research. Psychol Addict Behav 2017 Feb;31(1):1-20. [doi: 10.1037/adb0000237] [Medline: 28165272]
20. Powell BJ, Waltz TJ, Chinman MJ, Damschroder LJ, Smith JL, Matthieu MM, et al. A refined compilation of implementation strategies: results from the Expert Recommendations for Implementing Change (ERIC) project. Implement Sci 2015 Feb 12;10(1):21 [FREE Full text] [doi: 10.1186/s13012-015-0209-1] [Medline: 25889199]

Edited by A Mavragani; submitted 02.02.22; peer-reviewed by A Quanbeck, O Ogundaini; comments to author 21.04.22; revised version received 29.04.22; accepted 16.05.22; published 20.06.22.

Please cite as:
Wray TB
Exploring Whether Addictions Counselors Recommend That Their Patients Use Websites, Smartphone Apps, or Other Digital Health Tools to Help Them in Their Recovery: Web-Based Survey
JMIR Form Res 2022;6(6):e37008
URL: https://formative.jmir.org/2022/6/e37008
doi:10.2196/37008
PMID:35723917

Original Paper

Predicting Mental Health Status in Remote and Rural Farming Communities: Computational Analysis of Text-Based Counseling

Mark Antoniou[1], PhD; Dominique Estival[1], PhD; Christa Lam-Cassettari[2], PhD; Weicong Li[1], PhD; Anne Dwyer[1], PhD; Abìlio de Almeida Neto[3], PhD

[1]The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Penrith, Australia

[2]University of New South Wales, Kensington, Australia

[3]Centre for Work Health and Safety, NSW Government, Gosford, Australia

Corresponding Author:
Mark Antoniou, PhD
The MARCS Institute for Brain, Behaviour and Development
Western Sydney University
Locked Bag 1797
Penrith, 2751
Australia
Phone: 61 2 9772 6673
Email: m.antoniou@westernsydney.edu.au

Abstract

Background: Australians living in rural and remote areas are at elevated risk of mental health problems and must overcome barriers to help seeking, such as poor access, stigma, and entrenched stoicism. e-Mental health services circumvent such barriers using technology, and text-based services are particularly well suited to clients concerned with privacy and self-presentation. They allow the client to reflect on the therapy session after it has ended as the chat log is stored on their device. The text also offers researchers an opportunity to analyze language use patterns and explore how these relate to mental health status.

Objective: In this project, we investigated whether computational linguistic techniques can be applied to text-based communications with the goal of identifying a client's mental health status.

Methods: Client-therapist text messages were analyzed using the Linguistic Inquiry and Word Count tool. We examined whether the resulting word counts related to the participants' presenting problems or their self-ratings of mental health at the completion of counseling.

Results: The results confirmed that word use patterns could be used to differentiate whether a client had one of the top 3 presenting problems (depression, anxiety, or stress) and, prospectively, to predict their self-rated mental health after counseling had been completed.

Conclusions: These findings suggest that language use patterns are useful for both researchers and clinicians trying to identify individuals at risk of mental health problems, with potential applications in screening and targeted intervention.

(JMIR Form Res 2022;6(6):e33036) doi:10.2196/33036

KEYWORDS

e-mental health; text-based; counseling; Linguistic Inquiry and Word Count; LIWC; depression; anxiety; stress

Introduction

Rural Australians Are at Increased Risk of Mental Health Problems

Australians who live in rural and remote communities are at increased risk of adverse health outcomes because they face a combination of chronic, but unpredictable, stressors. Rural communities are those geographic areas located outside towns and cities, and remote areas are places that are isolated or considerably secluded from civilization. Overall, there are fewer employment opportunities than in urban centers, and reliance on primary industries leaves rural areas prone to financial instability owing to fluctuations in weather conditions, natural disasters (eg, drought, bushfires, floods, and cyclones), commodity and fuel prices, and currency exchange rates [1]. Remoteness increases the risk of mental illness, self-harm, and suicide [2]. Relative to urban areas, the suicide rate is 40%

higher in rural areas, increasing to 100% higher in remote areas. Suicide rates in both rural and remote areas are also increasing faster than in capital cities (between 2012 and 2016, suicide rates increased by 9.2% outside capital cities, compared with 2% in capital cities) [3]. In particular, farmers are more likely to commit suicide than other occupations [4-6]. Those most at risk fit the following profile: most of them are men (>90%), are young (mean suicide age of 41 years), have recently separated or divorced (20%), live alone (33%), are more likely to be farm laborers than farm owners or managers, and have a precipitating mental condition [7].

Barriers to Accessing Mental Health Services

Fewer mental health professionals work in rural and remote areas than in urban areas (70%-80% less than those in major cities). Those living in rural communities may be reluctant to seek counseling services owing to stigma, community gossip, entrenched stoicism, and views that help seeking is a sign of weakness [8,9].

In recent decades, technological advances have led to the implementation of e-mental health services that can circumvent some of the barriers mentioned previously and make valuable contribution to service delivery in rural areas [10]. Various technology-based approaches have been developed, including text-based, audio-delivered, or even audio-visual counseling and web-based counseling services that offer a suite of delivery methods (eg, Talkspace [11,12]). Text-based counseling may be particularly well suited to individuals who are reluctant to seek help owing to stigma and stoicism [13]. Synchronous text-based counseling involves the simultaneous participation of 2 parties (eg, a client and a therapist), who engage in real-time communication (eg, via text-based messaging). In all, two aspects of text-based counseling are likely to appeal to farmers living in rural and remote communities: (1) the anonymous nature of text-based interactions and (2) low-bandwidth delivery across great distances, which eliminates the need for transport or high levels of internet connectivity.

Evidence suggests that people generally disclose emotions in similar ways when using technology and face-to-face communication [14] and that text-based communication enables some people to better express their true-self qualities [15] or to disclose more personally confronting topics [16].

A consistent observation across studies is that text-based counseling takes longer than phone counseling [17,18] and generates fewer words than verbal exchanges [19]. Some patients expressed negative views about text-based communication related to less fluid interactions, reduced content covered, and impatience while waiting for the therapist to respond. Clients who prefer face-to-face or phone conversations have described text-based communication as too distant or impersonal. However, others have found the time delays created space to think, reflect, and communicate feelings without being disrupted by further questioning, as might occur in face-to-face sessions [20]. Those who viewed text-based counseling positively appreciated the distance, anonymity, security, privacy, and control over self-presentation [21].

The Effectiveness of Text-Based Counseling

Text-based counseling has been shown to be as effective as traditional face-to-face counseling for a variety of conditions including depression [12,22], anxiety [12,23], and emotional problems [17,21]. Therapist-guided internet-delivered treatments are effective in treating a range of mental health conditions and can be as effective as face-to-face treatments [24-26].

Text-based delivery has been rated as better than or equal to face-to-face therapy in several dimensions, including convenience, effectiveness, making progress with problems, and having access to help when needed [11]. However, it should be noted that text-based counseling may not be the optimal mode of service delivery for all clients [18] and that those who do not engage during text-based therapy will not show clinical improvement [27,28].

Computational Linguistic Analyses of the Text-Based Counseling

There is increasing evidence that analysis of e-mental health communications can be used to draw reliable inferences that can guide treatment. For instance, voice analysis systems have been developed that use artificial intelligence to improve treatment outcomes (eg, Eleos) or convey empathy and predict treatment engagement (eg, Lyssn AI). Text-based counseling is particularly attractive for subsequent computational linguistic analyses because it automatically documents the exchanges during the therapeutic process, thus avoiding the costs and difficulties associated with generating transcripts of audio-recorded sessions. This text chat offers possibilities that are not available to other service delivery methods: it permits the client and therapist to reread and reflect on their communication after the session has ended and it also opens up the research possibility of conducting analyses on text chats to identify linguistic patterns. Emerging literature suggest that language use patterns are reliable predictors of mental health status, but few studies have linguistically analyzed texts from individuals at risk of mental health problems. Most have mined social media [29-33] or web-based forums [34-37] to identify linguistic patterns that might be predictive of mental health status.

This study has been enabled by the development of computational linguistic tools, of which the most widely used is the Linguistic Inquiry and Word Count (LIWC) [38,39]. In addition to categorizing and counting words, LIWC offers an overview of the statistical distribution of words within predefined and psychologically meaningful categories. The capabilities of LIWC (and other similar programs or algorithms) have led researchers to explore the language use patterns of individuals with depression and other mental health conditions. Numerous studies have shown that increased use of first-person singular pronouns (eg, *I*, *me*, *my*, and *mine*) is indicative of depression [40-44], severity of depression and anxiety [45,46], general proneness to distress or negative emotionality [47,48], and suicidal ideation [49]. These promising findings suggest that language use patterns could conceivably serve as predictors of mental health, with potentially clinically significant applications.

Very few studies have analyzed text-based counseling. Nevertheless, the results from this small number of studies suggest that this may be a potentially fruitful avenue for future studies. Dirkse et al [50] found that greater use of negative emotion, anxiety, and sadness words positively correlated with heightened anxiety; greater use of negative emotion, sadness, and anger words positively correlated with heightened depression; and greater use of negative emotion and anger words positively correlated with heightened panic. Compatible findings showed that the use of negative emotion words predicted symptom improvement in outpatients being treated for personality disorders [51], and use of discrepancy words (eg, *would*, *should*, *wish*, and *hope*) reliably predicted depression improvement [52]. Patients with depression who used positive emotion words early in treatment tended to have good treatment outcomes, whereas the use of past focus words was associated with poor treatment outcomes [53]. These results suggest that word use may be used to determine an individual's psychological condition and future prognosis.

Owen et al [54] examined word use in a support intervention for women with early-stage breast cancer. More frequent use of words expressing anxiety and sadness (but not anger) was significantly correlated with improved emotional well-being at follow-up, whereas greater expression of sadness (but not anxiety or anger) was associated with improved quality of life.

Although this is an emerging research area, the level of sophistication in the analyses is rapidly improving. Seabrook et al [55] were able to reliably predict depression severity using negative emotion word instability, and interestingly, they created an emoji and internet slang supplement to the LIWC dictionary, which increased the accuracy of depression identification. In addition, it may soon be possible to combine demographic, linguistic, behavioral, and social data to construct sophisticated models to identify at-risk individuals [56].

This Study

This study examined an initiative funded by the Australian Government that provided text-based counseling to Australians in rural and remote communities through the *Virtual Psychologist* service [57]. The major aim of this study was to demonstrate the feasibility of using linguistic patterns in text-based counseling chats to predict whether an individual is experiencing depression, anxiety, or stress. The analysis was conducted using the LIWC software tool [38] on the text obtained from counseling sessions conducted over a 1-year period. The study also investigated whether language patterns were predictive of self-rated mental status. An additional consideration was to examine the characteristics of individuals who are using text-based counseling in rural and remote Australia.

Methods

Ethics Approval

This study was conducted in full compliance with the National Statement on Ethical Conduct in Human Research and approved by the Western Sydney University Human Ethics Committee (approval number H13309).

Recruitment

Participants were 320 clients of the Virtual Psychologist text-based counseling service, who used the service between August 2019 and September 2020 for ≥1 sessions. Virtual Psychologist is a privately owned, for-profit organization that offers counseling over a range of platforms (eg, text, voice, or video). This study involved only those clients who engaged in live text-based counseling. On average, each session lasted for 52 (SD 16) minutes. All the therapists in the Virtual Psychologist service were qualified psychologists. Funded by the Australian government, the text-based counseling service has been provided free of charge to any Australian farmer who feels that they need such service. The Virtual Psychologist service has not been scientifically evaluated, although it uses evidence-based counseling approaches. The Virtual Psychologist service was advertised through various platforms, such as radio, television, and social media (eg, Twitter and Facebook). Some participants were also referred by friends, family members, volunteers, police, or physicians. Participants were able to have as many sessions as they wanted and could ask for a session when they felt they needed it. All sessions were initiated by the participant. In most instances, the participant ended the session, usually once they felt that their pressing concerns had been addressed. All participants lived in rural or remote communities across Australia. Data from participants aged <18 years were excluded from the study.

Materials

Virtual Psychologist provided the data on a monthly basis to the researchers. The data consisted of the text from chat sessions, with metadata providing the date and time of each interaction and demographic information about the participant.

The LIWC tool [38] is the most widely used corpus of dictionaries for computational linguistic analyses of text data. It is a software program containing algorithms that enable it to count words belonging to different categories. To achieve this, LIWC compares words within an input text file with those within its dictionary. The output provides an overview of the statistical distribution of words within a text into predefined categories, including function words, pronouns, impersonal pronouns, verbs, auxiliary verbs, and past-tense words.

The LIWC dictionaries were customized to suit the Australian data set. To achieve this, Australian spellings were added to the standard LIWC American spellings (eg, Australian *agonise vs* US *agonize*), and, where necessary, equivalent Australian words were also added (eg, Australian *mobile vs* US *cellphone*). The Australianized dictionary is described in Multimedia Appendix 1.

Procedure

Participants were provided with the service's terms and conditions at the first point of contact with the *Virtual Psychologist* service. Then, they completed a short demographic survey, which was deidentified. Text from counseling sessions between participants and therapists were also deidentified (ie, names of people, workplaces, and landmarks were removed). Immediately after each session, participants received an SMS with a link to a short client survey regarding their mental health

presenting problems (choice from a list of 20 common presenting problems; Table 1) and their experience with the *Virtual Psychologist* service. The survey also contained the single-item self-rating of mental health, "How would you rate your mental health now?" (adapted from the study by Althoff et al [58]) on a 5-point scale ranging from *poor* to *excellent*. The survey was optional for participants, and the response rate for the self-rating of mental health was found to be relatively low. In July 2020, the participants who had not completed the survey in their sessions were contacted and asked to respond to the self-rating.

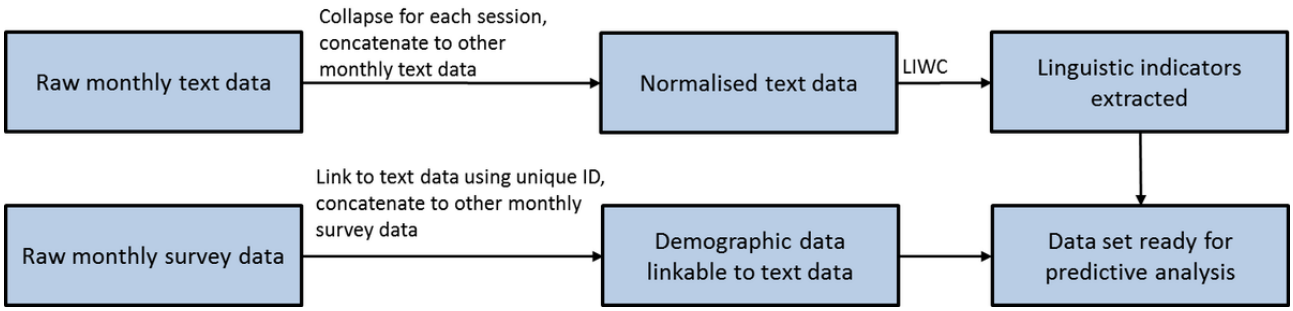
Table 1. Distribution of the number of text-based counseling sessions completed by each of the 270 participants. A total of 94.8% (256/270) of participants completed between 1 and 7 sessions (N=270).

Number of sessions	Participants, n (%)
1	98 (36.3)
2	61 (22.6)
3	32 (11.9)
4	34 (12.6)
5	15 (5.6)
6	12 (4.4)
7	4 (1.5)
8	3 (1.1)
9	3 (1.1)
10	2 (0.7)
11	1 (0.4)
12	2 (0.7)
13	2 (0.7)
14	1 (0.4)

Data Processing

Owing to privacy reasons, names, locations, and other identifying information in the raw text data were manually identified and replaced using labels such as NAME or PLACE by Virtual Psychologist before the data were provided to the research team. Each participant and chat session were assigned a unique participant ID and session ID, respectively. For each session, the chat text from the participant was first aggregated. To ensure compatibility with LIWC, the aggregated text data were preprocessed (normalized and restructured) using the MATLAB software developed by MathWorks. Normalization involves common text cleaning operations, including removing punctuations, extra spaces, and returns and converting all words to lowercase. Emojis and other internet slangs (if any) were left unchanged, together with spelling errors. Chat sessions containing <30 words from participants were removed; this was usually owing to participants being unable to continue soon after initiating the session (eg, poor cellular network coverage or work-related or personal situation requiring them to leave the session). A total of 33.02% (381/1154) of the sessions for 15.6% (50/320) of the participants met this criterion and were excluded from the analysis. The final data set comprised 66.98% (773/1154) of the sessions, involving 84.4% (270/320) of the participants. Figure 1 shows the workflow for data processing and analyses.

Figure 1. Workflow for the data processing and analyses. LIWC: Linguistic Inquiry and Word Count.



Data Analysis

Linguistic indicators were extracted for each session using LIWC [38,39]. We explored the relationships between the linguistic patterns in the chat sessions and participants' self-reported psychological concerns at service entry and their mental well-being self-rating after receiving counseling. Of all possible linguistic indicators, the following 21 indicators were

selected as independent variables (predictors) to be used in the predictive analysis: word count, analytical thinking score, clout score, authenticity score, emotional tone, first-person singular pronouns, positive emotions, negative emotions, causation, insight, discrepancy, social processes, functional words, other words, affect, cognitive processes, drives, personal concerns, past focus, present focus, and future focus. The selection of these 21 indicators was based on the following reasons: first, they cover most of the word categories available in LIWC; second, the rest of the indicators available in LIWC, such as punctuation marks and relativity (motion, space, and time), were considered to have little relevance to self-reported presenting problems and self-rated mental well-being, and thus were excluded; and third, previous studies [40-54] showed that some mental health problems, such as anxiety and depression, were highly correlated with a selection of these indicators. Participants' self-reported mental health problems at service initiation and self-rated mental well-being from the survey were treated as dependent variables in the predictive analysis. Quadratic discriminant analyses with 5-fold cross validation were conducted in MATLAB to explore the relationship between the independent and dependent variables. Each discriminant model uses a different combination of the 21 linguistic indicators as predictors to calculate the probabilities of classification response and, then, outputs the predicted classification label based on the highest probability. The performance of each model was compared to determine which discriminant model performed optimally using 3 metrics: Area under the Receiver Operating Characteristic curve (AUC), general prediction accuracy, and average prediction accuracy when the prediction probability was set to 70%, 80%, and 90%. Models with highest AUC and accuracy and lowest number of predictors were preferred.

Results

Sample Description

The text data used in the analyses reported here were collected between August 2019 and September 2020. They consist of 100% (1154/1154) of the text-based counseling sessions from 100% (320/320) of the participants who engaged with the *Virtual Psychologist* service. Following data cleaning and preprocessing, the final data set for linguistic and predictive analysis comprised 66.98% (773/1154) of the sessions from 84.4% (270/320) of the participants. The distribution of sessions per participant is shown in Table 1.

Characteristics of Individuals Who Engaged in Text-Based Counseling

On average, participants completed 3.6 (SD 3.2) sessions of text-based counseling; however, there was considerable variability in the number of sessions completed. Most participants (256/270, 94.8%) engaged in 1 to 7 sessions; however, some engaged in as many as 14 sessions. Approximately one-third (98/270, 36.3%) of the participants engaged in only 1 session. For each session, the client sent an average of 11 (SD 11; range 1-84) messages. The total number of words per session also varied widely, with an average session containing 357 (SD 300) words exchanged between the therapist and the client. This is consistent with the literature reporting that text-based chat is slower than verbal communication and results in fewer words being exchanged between conversational partners [17,18].

Data collection commenced in August 2019. However, as shown in Table 2, the number of monthly sessions increased from March 2020 owing to increase in the number of participants using the counseling service. This was likely precipitated by 2 events. First, Australia experienced unprecedented bushfires in late 2019, extending into early 2020, and rural areas were the most badly affected. Second, the first confirmed case of COVID-19 in Australia was identified on January 25, 2020 [59], resulting in Australian borders being closed to nonresidents on March 20, 2020, and government restrictions (social distancing rules and closing of nonessential services) being put in place on March 21, 2020. In our text-based chat data, bushfire-related events were mentioned 1 to 9 times per month, and COVID-19 was mentioned 2 to 22 times per month during the period from December 2019 to September 2020.

Table 2. Number of text-based counseling sessions completed by participants in each month from August 2019 to September 2020 (N=773).

Month and year	Sessions, n (%)
August 2019	45 (5.8)
September 2019	39 (5)
October 2019	33 (4.3)
November 2019	46 (5.9)
December 2019	43 (5.6)
January 2020	26 (3.4)
February 2020	18 (2.3)
March 2020	54 (6.9)
April 2020	99 (12.8)
May 2020	76 (9.8)
June 2020	56 (7.2)
July 2020	73 (9.4)
August 2020	97 (12.5)
September 2020	68 (8.8)

Regarding sex, most participants were women (167/270, 61.9%). As shown in Table 3, women outnumbered men by 3:1 and also completed more sessions. Of the 270 participants, 44 (16.3%) participants did not disclose their sex.

Table 3. Number of sessions and number of participants for each sex.

Sex	Sessions (N=773), n (%)	Participants (N=270), n (%)
Female	500 (64.7)	167 (61.9)
Male	162 (21)	59 (21.9)
Undisclosed	111 (14.3)	44 (16.3)

The age distribution of the participants is shown in Table 4. The sample primarily comprised young adults (age ranges 18-21 years and 22-29 years). This is consistent with reports that individuals who are comfortable with using technology are more likely to engage in text-based counseling. However, interestingly, the discrepancy between participants aged 18 to 21 years and 30 to 40 years decreases when we inspect the number of sessions completed, suggesting that the average participants aged 30 to 40 years engaged in a greater number of counselling sessions than the average participants aged 18 to 21 years. This is encouraging, as men aged 41 years are most at risk of serious mental health problems and suicide [60]. Older adults were fewer in number and, on average, engaged in fewer sessions than their younger counterparts (2 sessions).

Table 4. Number of sessions completed by each age category and participants' age distribution.

Age categories (years)	Sessions (N=773), n (%)	Participants (N=270), n (%)
18-21	196 (25.4)	74 (27.4)
22-29	275 (35.6)	77 (28.5)
30-40	133 17.2)	43 (15.9)
41-50	78 (10.1)	39 (14.4)
51-60	36 (4.7)	15 (5.6)
61-70	29 (3.8)	12 (4.4)
Undisclosed	26 (3.4)	10 (3.7)

Upon referral to the *Virtual Psychologist* counseling service, each participant's self-reported mental health concern was recorded (Table 5). The number of presenting problems reported by each participant ranged from 1 to 5, with an average value of 1.76 (SD 0.96). The top 3 mental health conditions that clients presented with were anxiety, depression, and stress, and they comprised approximately half of the total number of sessions. These 3 presenting problems sometimes overlapped for the same individual. Of the 270 participants, 26 (9.6%) participants reported having both depression and anxiety, 10 (3.7%) reported both depression and stress, 9 (3.3%) reported both anxiety and stress, and 4 (1.5%) reported having all the 3 problems.

Approximately one-fourth of all sessions fell into the *other* and *undisclosed* categories. Apart from the explanation that participants may not be able to find the right category for their problems, this could also suggest that even for an anonymous and privacy-focused method of e-mental health service provision, there remains a considerable number of individuals for whom disclosure, and presumably stigma, remains as an issue (even when it does not prevent seeking help).

Table 5. Presenting problems that led participants to seek counseling, expressed as distribution of the number of text-based counseling sessions completed[a].

Presenting problem	Sessions (N=773), n (%)	Participants (N=270), n (%)
Anxiety	152 (19.7)	71 (26.3)
Depression	143 (18.5)	79 (29.3)
Stress	61 (7.9)	34 (12.6)
Family issues	50 (6.5)	30 (11.1)
Relationship issues	49 (6.3)	34 (12.6)
Grief and loss	25 (3.2)	12 (4.4)
Trauma issues	15 (1.9)	11 (4.1)
Suicidal thoughts	13 (1.7)	10 (3.7)
Anger	6 (0.8)	6 (2.2)
Work problems	6 (0.8)	5 (1.9)
Domestic violence	4 (0.5)	4 (1.5)
Isolation or loneliness	4 (0.5)	2 (0.7)
Critical incident	3 (0.4)	3 (1.1)
Self-harm	3 (0.4)	3 (1.1)
COVID-19	2 (0.2)	2 (0.7)
Eating disorders	1 (0.1)	1 (0.4)
Friend issues	1 (0.1)	1 (0.4)
Health concerns	1 (0.1)	1 (0.4)
LGBTI[b] issues	1 (0.1)	1 (0.4)
Physical abuse	1 (0.1)	1 (0.4)

[a]Technical issues and undisclosed presenting problems (232/773, 30% of the sessions for 121/270, 44.8% of the participants) are not listed.

[b]LGBTI: lesbian, gay, bisexual, transgender, and intersex.

Of the 773 completed sessions, 165 (21.3%) responses obtained from 38.9% (105/270) of the participants were recorded for the single-item self-rating of mental well-being, "How would you rate your mental health now?" on a 5-point scale ranging from *poor* to *excellent*. A total of 29.7% (49/165) of responses from 40.9% (43/105) of the participants were collected immediately after the counseling sessions, and the remaining responses were collected in July 2020. Of the 21.3% (165/773) sessions with responses to the self-rating question, 51.5% (85/165) reported anxiety, depression, and stress as presenting problems, whereas for those without responses to the self-rating question, 44.7% (272/608) reported these 3 presenting problems. Among those participants who responded to the self-rating question, 72.4% (76/105) were women, 15.2% (16/105) were men, and 12.4% (13/105) chose to remain undisclosed, whereas among those who did not respond, 57.6% (95/165) were women, 21.8% (36/165) were men, and 20.6% (34/165) chose to remain undisclosed. Regarding age distribution, of the participants who responded to the self-rating question, those in the age groups of 18 to 21 years, 21 to 29 years, and 30 to 40 years were 4% less than those who did not respond, but those in the age groups of 41 to 50 years, 51 to 60 years, and 61 to 70 years were 5.5% more than those who responded. Therefore, older women who reported anxiety, depression, and stress were more likely to respond to the self-rating question. As shown in Table 6, most participants chose *fair*, although responses varied widely, and made use of the full range of response options available. The average self-rating score was 2.7 (SD 1.3).

Table 6. Participant responses to the single-item self-rating of mental well-being, "How would you rate your mental health now?" on a 5-point scale ranging from poor to excellent (N=165).

Self-ratings of mental well-being	Sessions, n (%)
Poor	34 (20.6)
Fair	53 (32.1)
Good	30 (18.2)
Very good	29 (17.6)
Excellent	19 (11.5)

Linguistic Analysis

The 4 basic LIWC scores are shown in Table 7: analytical thinking, clout, authenticity, and emotional tone.

Table 7. Descriptive statistics of Linguistic Inquiry and Word Count scores for the basic summary variables: analytical thinking, clout, authenticity, and emotional tone. Scores are calculated based on the text from each session. Summary variable scores range from 1 to 99.

Indicators	Score, mean (SD)	Score, median (range)
Analytical thinking	24 (19)	19 (1-95)
Clout	34 (26)	28 (1-99)
Authenticity	75 (26)	86 (1-99)
Emotional tone	57 (34)	60 (1-99)

The distributions of these 4 basic LIWC dimensions are shown in Table 8. Analytical thinking scores showed a shallow positive skew, with most of those scores falling within the range of 0 to 50. This indicates that participants were using a language style similar to a narrative, focused on their personal experiences. High analytical thinking scores are associated with better academic performance in tertiary education [61]. The observed concentration of scores on the other half of the scale appears to be a valid representation of the sample population being studied, that is, farmers living in rural areas.

Table 8. Distribution of Linguistic Inquiry and Word Count scores for the basic dimensions—analytical thinking, clout, authenticity, and emotional tone (N=773).

Score (range)	Analytical thinking, n (%)	Clout, n (%)	Authenticity, n (%)	Emotional tone, n (%)
0-10	180 (23.3)	134 (17.3)	23 (3)	86 (11.1)
11-20	231 (29.9)	165 (21.3)	21 (2.7)	64 (8.3)
21-30	142 (18.4)	114 (14.7)	32 (4.1)	72 (9.3)
31-40	84 (10.9)	93 (12)	26 (3.4)	57 (7.4)
41-50	50 (6.5)	59 (7.6)	40 (5.2)	55 (7.1)
51-60	32 (4.1)	68 (8.8)	37 (4.8)	53 (6.8)
61-70	24 (3.1)	45 (5.8)	60 (7.8)	50 (6.5)
71-80	18 (2.3)	32 (4.1)	86 (11.1)	73 (9.4)
81-90	10 (1.3)	24 (3.1)	132 (17.1)	61 (6.6)
91-100	2 (0.2)	39 (5)	316(40.9)	212 (27.4)

Clout refers to social status, confidence, or leadership [62]. Clout scores showed a shallow positive distribution and were somewhat more evenly distributed across the range of scores. This may reflect different ranks or responsibilities within the sample, such as farm laborers versus farm managers and owners.

Authenticity scores showed a sharp negative distribution. Higher authenticity scores indicate truthfulness, humility, and vulnerability [63,64]. Encouragingly, this indicates that most participants were using language associated with being truthful. This is consistent with literature suggesting that text-based counseling offers high degree of privacy and anonymity, giving users time and space to select the right words to express themselves [20,21] and reveal more truthful information [16].

Emotional tone scores were the most evenly distributed among the 4 LIWC summary variables. High scores (>50) reflect more positive emotional tone [65]. Participants spanned the full range of the scale, and the mean was 57, indicating a neutral to positive emotional tone. There was a gradual increase in the number of scores toward the negative end of the scale, indicating that the sample contained individuals experiencing severe negative

emotions. Encouragingly, there was also a sharp spike in the most positive interval of the scale (90-99), indicating that many participants were using positive emotion words, which included positive feelings or expressions of gratitude to the therapist.

The category with the best-established relationship to mental health is that of first-person singular pronouns. Overuse of first-person singular pronouns (eg, *I*, *me*, *my*, and *mine*) is a marker of depression [40,42,44] and predicts the severity of depressive symptoms 8 months after treatment [46]; however, recent findings suggest that first-person singular pronoun use may be indicative of general proneness to distress or negative emotions rather than of depression specifically [47,48]. As shown in Table 9, use of first-person singular pronouns comprised 10% of the words.

Table 9. Percentage of words falling within the Linguistic Inquiry and Word Count categories: first-person singular pronouns, positive emotion, negative emotion, causation, discrepancy, insight, and social processes. The indicators are calculated based on the text from each session.

Indicators	Words (%), mean (SD)	Words (%), median (range)
First-person singular pronouns	10 (3.3)	10.3 (0-22.5)
Positive emotion	5.3 (3.2)	4.4 (0-27.8)
Negative emotion	2.7 (1.7)	2.6 (0-10.1)
Causation	1.6 (1.1)	1.6 (0-6.5)
Insight	2.8 (1.7)	2.8 (0-10.3)
Discrepancy	2 (1.5)	1.8 (0-12.5)
Social processes	10.2 (4.9)	9.5 (0-35.3)

Social process words (eg, *share* and *we*) also comprised approximately 10% of the words within a session. This is to be expected, as the participants were reflecting on their relationships with others.

Use of positive (eg, *happy* and *brave*) and negative (eg, *sad* and *desperate*) emotion words is known to relate to mental health and symptom severity. Individuals with depression use more negative and fewer positive emotion words [41,43]. Reduced use of negative emotion words predicts symptom improvement [51]. In this study, the frequency of positive emotion words shows a positive skew, as may be expected for individuals undergoing psychological counseling.

There is some evidence that the use of absolutist words (ie, words that indicate certainty such as *always*, *totally*, *constantly*, *forever*, *completely*, and *entire*) may predict suicidal ideation better than negative emotion words or first-person pronouns [34,36]. We expect those participants experiencing greater psychological distress to make greater use of causation words [50,51] and less use of discrepancy words [52]. As shown in Table 10, the distributions of causation words (eg, *because*, *aggravate*, and *basis*) resemble that of the negative emotion and insight categories. The distribution of discrepancy words (eg, *would not*, *unusual*, *abnormal*, and *impossible*) shown in Table 10, is moderately more positively skewed, which suggests that there is considerable variability within the data regarding the severity of mental health issues being experienced by participants.

Table 10. Distribution of words per session for the Linguistic Inquiry and Word Count categories—first-person singular pronouns, positive emotion words, negative emotion words, causation words, insight words, discrepancy words, and social processes.

Indicators	Words (%), range; %									
First-person singular pronouns	0-2.3; 1.9	2.3-4.6; 4.9	4.6-6.9; 10.5	6.9-9.2; 19.4	9.2-11.5; 29.1	11.5-13.8; 24.5	13.8-16.1; 8.3	16.1-18.4; 1	18.4-20.7; 0.3	20.7-23; 0.1
Positive emotion	0-2.8; 16	2.8-5.6; 51	5.6-8.4; 19.7	8.4-11.2; 7.6	11.2-14; 3.6	14-16.8; 1.4	16.8-19.6; 0.1	19.6-22.4; 0.3	22.4-25.2; 0.1	25.2-28; 0.1
Negative emotion	0-1.1; 14.9	1.1-2.2; 23.5	2.2-3.3; 28.8	3.3-4.4; 17.6	4.4-5.5; 8.5	5.5-6.6; 4.4	6.6-7.7; 1.8	7.7-8.8; 0.1	8.8-9.9; 0	9.9-11; 0.3
Causation	0-0.65; 21	0.65-1.3; 15.4	1.3-1.95; 27.8	1.95-2.6; 21.2	2.6-3.25; 9.8	3.25-3.9; 2.7	3.9-4.55; 0.9	4.55-5.2; 0.5	5.2-5.85; 0.3	5.85-6.5; 0.4
Insight	0-1.1; 16.2	1.1-2.2; 20.6	2.2-3.3; 27.2	3.3-4.4; 21.3	4.4-5.5; 9.3	5.5-6.6; 4	6.6-7.7; 0.9	7.7-8.8; 0.1	8.8-9.9; 0.3	9.9-11; 0.1
Discrepancy	0-1.3; 29.2	1.3-2.6; 45	2.6-3.9; 17.2	3.9-5.2; 4.8	5.2-6.5; 1.8	6.5-7.8; 1.2	7.8-9.1; 0.5	9.1-10.4; 0.1	10.4-11.7; 0	11.7-13; 0.1
Social processes	0-3.6; 4.9	3.6-7.2; 23.4	7.2-10.8; 30.7	10.8-14.4; 23	14.4-18; 11.3	18-21.6; 4.4	21.6-25.2; 1.6	25.2-28.8; 0.3	28.8-32.4; 0.4	32.4-36; 0.1

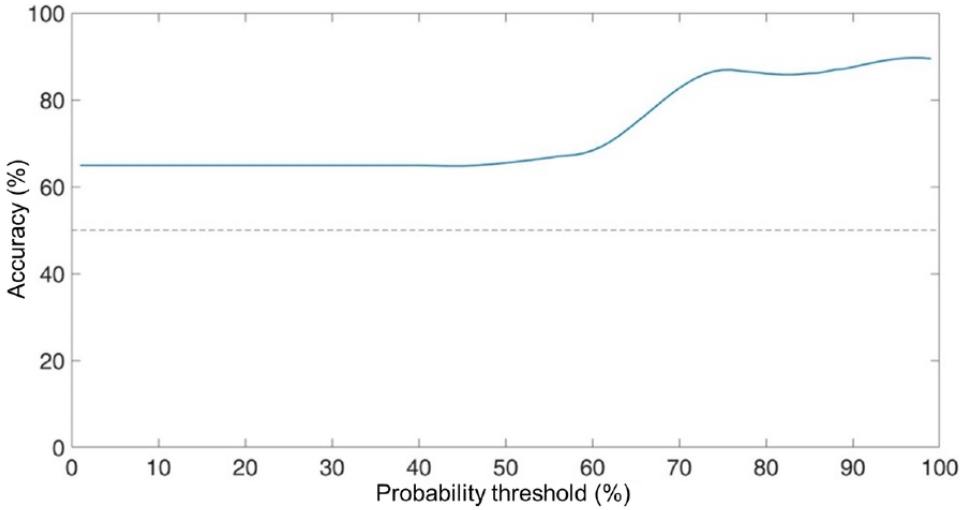
Predictive Analysis

Overview

Quadratic discriminant analyses were conducted to explore the relationship between the linguistic categories provided by LIWC and mental health status. Each discriminant model used a different combination of linguistic categories as predictors to calculate the probabilities of classification response and, then, outputs the predicted classification label based on highest probability. Then, 5-fold cross validation was applied to each of the discriminant models. The performance of the different models was compared to determine which discriminant model performed optimally. Model performance was assessed using the following three metrics:

1. Examining the AUC: Interpretation of AUC varies across disciplines. In applied psychology, given the large number of variables that can influence human behavior, AUC values ≥0.71 are considered as strong effects [66].
2. General prediction accuracy.
3. Average prediction accuracy when the prediction probability was set to 70%, 80%, and 90% (calculated from the accuracy curve; refer to example shown in Figure 2). The prediction accuracy increases considerably when the prediction probability is >70%.

Figure 2. Example of an accuracy curve showing prediction accuracy when the prediction probability is set at different thresholds. The dashed line shows the accuracy at chance level (50% for binary classification).



Binary Classification of Mental Health Presenting Problem

First, we examined whether language use patterns could be used to discriminate the top 3 presenting problems (ie, anxiety, depression, and stress; Table 5) from the remaining pool of presenting problems (binary classification). This distinction was deemed important because these 3 presenting problems were the most frequently occurring problems within the data set, and they are the most studied in the literature. Accurate, scalable classification would be useful for screening and for targeted interventions.

Before reporting the binary classification results, we inspect the differences in LIWC counts between the top 3 presenting problems and the pool of remaining presenting problems. To do this, we present boxplots for each LIWC count of interest. Boxplots are a standardized way of visualizing the distribution of data by presenting the median, first and third quartiles (edges of the box), and minimum and maximum (error bars); they indicate the spread of the data, whether it is symmetrical, how tightly it is grouped, and skewness. Differences in boxplots between classification options would be indicative of accurate predictions in the corresponding discriminant analyses. Figure 3 shows boxplots for the 4 basic LIWC counts. The top 3 presenting problems, relative to the others, were differentiated by lower clout and authenticity scores and higher emotional tone scores.

Figure 3. Boxplots of the 4 basic Linguistic Inquiry and Word Count counts (clout, authenticity, emotional tone, and analytical thinking) for the top 3 presenting problems (red) and the remaining pool of other presenting problems (blue).

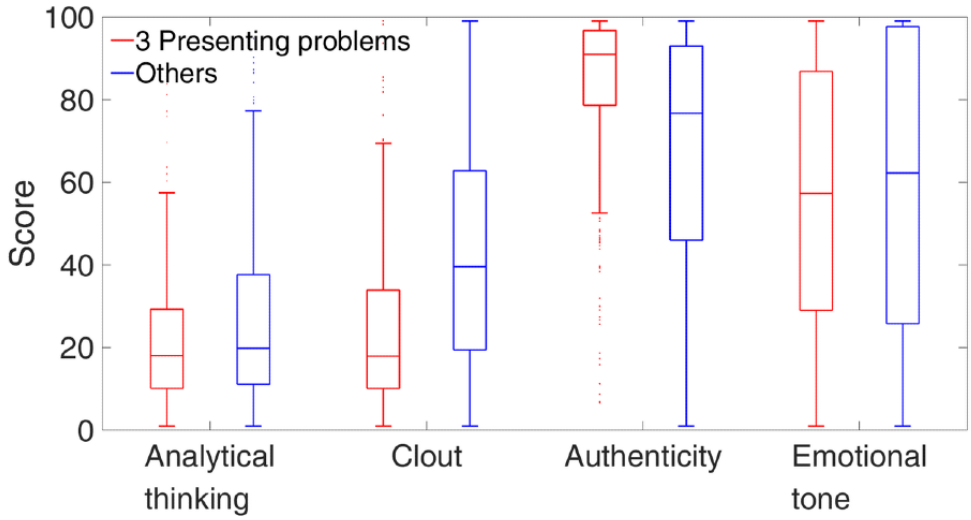


Figure 4 shows boxplots for the LIWC categories. In the upper panel, the top 3 presenting problems were differentiated from the others by an increased use of first-person singular pronouns and insight words, but less use of discrepancy and social process words. In the lower panel, the 2 classifications may also be differentiated by words in 3 predefined LIWC categories: *cognitive processes*, *future focus*, and *drives*.

Figure 4. Boxplots of the Linguistic Inquiry and Word Count categories for the top 3 presenting problems (red) and the remaining pool of other presenting problems (blue).

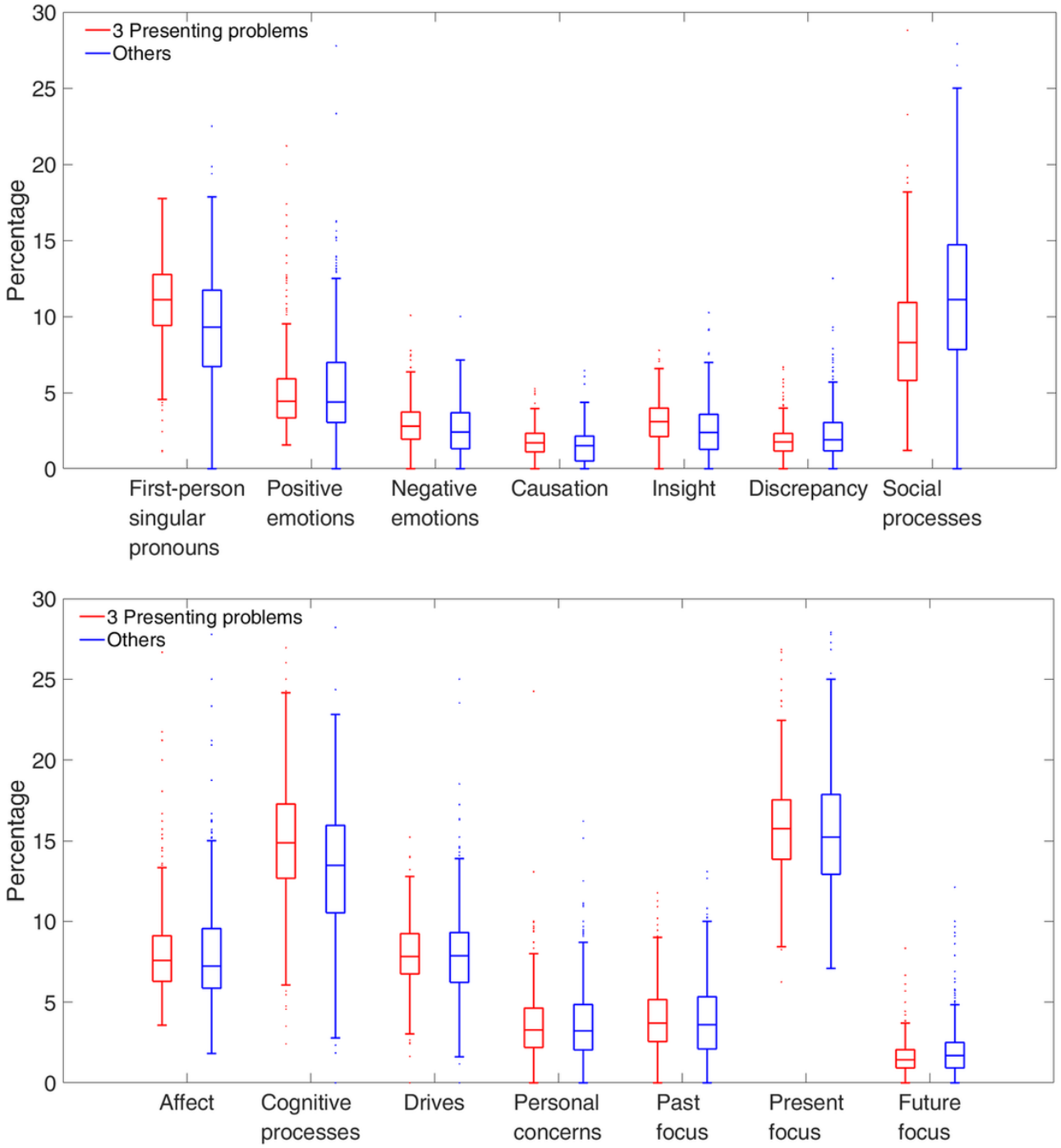


Table 11 presents the evaluation metrics (AUC, general accuracy, and average accuracy) for the best 5 discriminant models for discriminating the top 3 from the remaining pool of presenting problems using combinations of 1 to 5 and all predictors (binary classification). The data used in the binary classification comprised linguistic predictors extracted from 66.98% (773/1154) of the sessions with 84.4% (270/320) of the participants. Table 11 shows that general accuracy of the best discriminant models reached approximately 70% (exceeding the 50% chance level), with AUC of 0.76, indicating good discrimination and a strong effect. When the confidence of prediction was high, the average accuracy of prediction reached approximately 80%. Interestingly, increasing the number of predictors in the discriminant models did not offer continuous improvements in any of the evaluation metrics. Regarding the trade-off between model performance (here, accuracy and AUC) and model size (ie, the number of predictors used), models with 3 to 4 predictors seem to be optimal in terms of the balance between good accuracy and model complexity. Regarding LIWC categories, clout, future focus, discrepancy, emotional tones, drives, social processes, insight, and first-person singular pronouns were the most frequently occurring predictors among the discriminant models listed in Table 11.

Table 11. Best 5 models for discriminating the top 3 from the remaining pool of presenting problems.

Number of predictors and predictor names	AUC[a]	General accuracy (%)	Average accuracy[b] (%)	F1 score
1				
Clout score	0.70	64.9	86	0.68
Social processes	0.68	62.2	84.2	0.64
Authenticity score	0.66	61.4	81.5	0.67
First-person singular pronouns	0.66	62.7	84	0.65
Word count	0.64	56.8	N/A[c]	0.45
2				
Clout score+discrepancy	0.74	66.9	80.4	0.70
Clout score+functional	0.73	67.3	82.5	0.70
Clout score+future focus	0.73	67	81	0.70
Clout score+drives	0.73	66	82.9	0.68
Clout score+insight	0.73	68.2	83.6	0.69
3				
Clout score+discrepancy+focus future	0.75	67.1	77.6	0.70
Clout score+drives+future focus	0.75	67.8	79.4	0.70
Insight+social processes+functional	0.74	67.8	81.8	0.69
Clout score+authenticity score+future focus	0.74	67	78	0.69
Clout score+insight+drives	0.74	68.3	81.5	0.70
4				
Clout score+positive emotions+discrepancy+future focus	0.76	67.9	77.7	0.71
Clout score+emotional tone score+discrepancy+future focus	0.76	66.9	78.6	0.70
First-person singular pronouns+discrepancy+social processes+future focus	0.75	67.7	77.4	0.70
Clout+first-person singular pronouns+discrepancy+future focus	0.75	68.3	77.8	0.71
Clout+insight+drives+future focus	0.75	68.4	79.1	0.70
5				
Clout score+emotional tone score+discrepancy+functional+future focus	0.76	68.8	76.5	0.70
Clout score+emotional tone score+discrepancy+drives+future focus	0.76	67	77.3	0.71
Clout score+emotional tone score+discrepancy+social processes+future focus	0.76	68.4	77.9	0.71
Clout score+emotional tone score+insight+discrepancy+future focus	0.76	68.3	77.8	0.71
Word count+clout score+emotional tone score+discrepancy+ffuture focus	0.76	67.7	78.5	0.70
2[d] All predictors	0.72	63.8	67.9	0.66

[a]AUC: Area under the Receiver Operating Characteristic curve.

[b]Average accuracy when the predicted probability threshold is set to 70%, 80%, and 90%.

[c]N/A: not applicable.

[d]The best models with 6-20 predictors (total of 75 items) have lower AUC, general accuracy, average accuracy, and F1 score, thus are omitted.

Multiclass Classification of Presenting Problem

The second set of results relates to the task of using language patterns to differentiate between the top 3 mental health presenting problems (ie, anxiety, depression, or stress). For this multiclass classification, chance level is 33.3%. Figure 5 shows boxplots for the 4 basic LIWC counts. For all the 4 counts, there was considerable overlap across the 3 presenting problems. Of the 4 counts, analytical thinking score was the most promising for differentiating anxiety from depression and stress (red is lower than black and blue).

Figure 5. Boxplots of the 4 basic Linguistic Inquiry and Word Counts (analytical thinking, clout, authenticity, and emotional tone) for the top 3 presenting problems: anxiety (red), depression (black), and stress (blue).

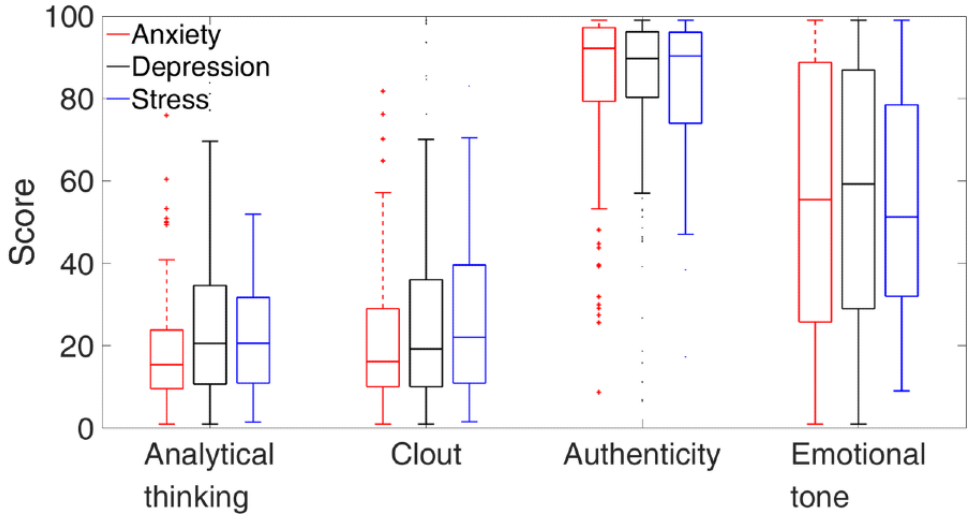


Figure 6 shows individual boxplots for the LIWC categories for each of the top 3 presenting problems. Again, there was considerable overlap across the top 3 presenting problems, suggesting that they share common features. Of the available LIWC categories, the most promising category in terms of differentiation was the first-person singular pronouns, which showed elevated counts for anxiety and depression relative to stress. However, the high degree of overlap across the LIWC categories suggests that it may be difficult to differentiate the top 3 presenting problems from one another.

Figure 6. Boxplots of the Linguistic Inquiry and Word Count categories for each of the top 3 presenting problems: anxiety (red), depression (black), and stress (blue).

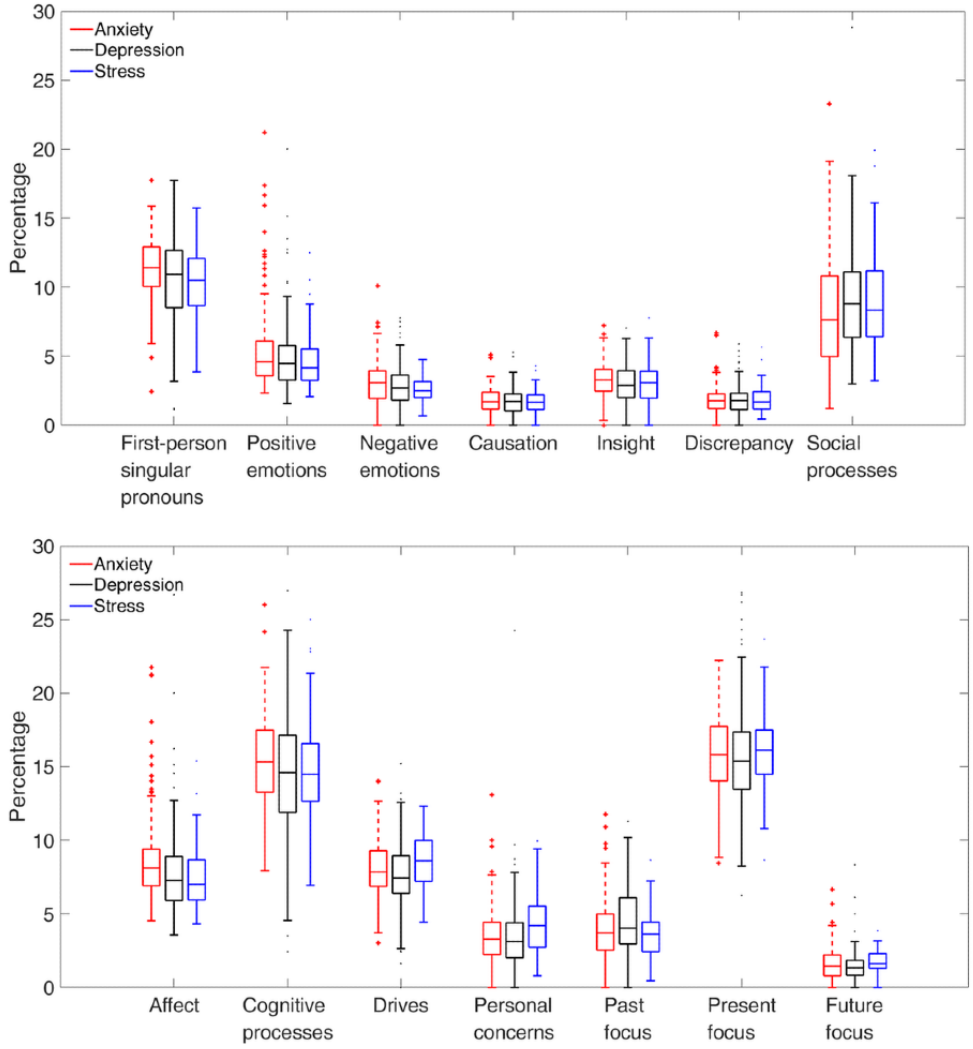


Table 12 presents the evaluation metrics for the best 5 discriminant models for discriminating between the top 3 mental health presenting problems (anxiety, depression, and stress) using combinations of 1 to 5 and all predictors. The data used in the classification comprised linguistic predictors extracted from 30.85% (356/1154) of the sessions with 44.7% (143/320) of the participants. Table 12 shows that the general accuracy of the best discriminant models was slightly >50% (compared with 33.3% chance level). Cohen κ coefficient reached 0.21, showing fair agreement between prediction and ground truth. When the confidence of prediction was high, the average accuracy of prediction for most discriminant models were between 50% and 70%. Increasing the number of predictors did not substantially improve prediction accuracy. These results suggest that it is difficult to differentiate the top 3 presenting problems (ie, anxiety, depression, and stress) on the basis of LIWC categories, even though performance was well above chance level. Of the LIWC categories, analytical thinking score, cognitive processes, first-person singular pronouns, past focus, and present focus were the most frequently occurring predictors among the discriminant models listed in Table 12.

Table 12. Best 5 discriminant models for differentiating between the top 3 mental health presenting problems (anxiety, depression, and stress).

Number of predictors and predictor names	Cohen κ coefficient	General accuracy (%)	Average accuracy (%)
1			
Analytical thinking score	0.12	49.2	71.6
Cognitive processes	0.12	49.2	52.6
Affect	0.12	48.6	33.3
First-person singular pronouns	0.09	47.2	83.3
Social processes	0.09	46.9	N/A[a]
2			
Analytic thinking score+first-person singular pronouns	0.14	50.3	75.1
Analytical thinking score+past focus	0.14	50.3	74.2
Analytical thinking score+cognitive processes	0.14	50	70.6
First-person singular pronouns+cognitive processes	0.14	50	54.2
First-person singular pronouns+past focus	0.14	50	57.1
3			
First-person singular pronouns+negative emotions+functional	0.18	50.6	66.8
Analytical thinking score+past focus+present focus	0.17	52	71.7
Analytical thinking score+emotional tone score+drives	0.17	51.4	77.5
Emotional tone score+cognitive processes+drives	0.17	51.4	66.7
First-person singular pronouns+cognitive processes+present focus	0.17	51.4	56.2
4			
Negative emotions+social processes+functional+focus present	0.21	50.8	68.7
Negative emotions+affect+drives+past focus	0.21	48.9	60.6
Analytical thinking score+affect+cognitive processes+present focus	0.2	52	64.6
Word count+analytical thinking score+causation+past focus	0.19	52	66.1
Word count+analytical thinking score+causation+present focus	0.19	52	62.7
21[b] All predictors	0.09	43.3	46.3

[a]N/A: not applicable.

[b]The best models with 5-20 predictors (total of 80 items) have lower Cohen κ coefficient, general accuracy, and average accuracy, thus are omitted.

Binary Classification of Self-rating of Mental Well-being

The next set of analyses was regarding whether language use patterns can identify individuals with the poorest future mental health status by discriminating those individuals who rated their health as *poor* from the rest, that is, those who assigned the ratings *fair* to *excellent* (binary classification). Again, this distinction is important because those participants who rated their mental health as *poor* are more likely to require targeted intervention.

Discriminant models used all participants (105/105, 100%) and linguistic indicators extracted from 21.3% (165/773) of the corresponding chat sessions. Table 13 presents the evaluation metrics for the best 5 discriminant models for discriminating *poor* rating for self-rated mental health from other ratings

ranging from *fair* to *excellent* using combinations of 1 to 5 and all predictors in each model. Table 13 shows that the general accuracy of the best discriminant models reached approximately 80% (exceeding the 50% chance level), with AUC reaching 0.73, showing good discrimination. When the confidence of prediction was high, the average accuracy of prediction was within the range of 80% to 90%. Increasing the number of predictors did not substantially improve AUC and prediction accuracy. Again, regarding the trade-off between model performance (here, accuracy and AUC) and model size (ie, the number of predictors used), models with 4 to 5 predictors were optimal in terms of the balance between good accuracy and model complexity. Analytical thinking score, positive emotions, discrepancy, causation, drives, first-person singular pronouns, and cognitive processes were the most frequently occurring predictors among the best-performing discriminant models.

Table 13. Best 5 discriminant models for discriminating poor response to self-rated mental health from other ratings (fair to excellent).

Number of predictors and predictor names	AUC[a]	General accuracy (%)	Average accuracy (%)	F1 score
1				
Analytical thinking score	0.59	79.4	84.3	0.56
Discrepancy	0.59	79.4	82.7	0.56
Insight	0.53	77.6	53.5	0.49
Present focus	0.51	79.4	75.9	0.55
Causation	0.50	79.4	82	0.55
2				
Analytical thinking score+other words	0.66	79.4	88.4	0.55
Discrepancy+drives	0.64	79.4	87.2	0.55
Clout score+social processes	0.63	79.4	87.2	0.55
Analytical thinking score+emotional tone score	0.62	79.4	85.9	0.56
Analytical thinking score+positive emotions	0.62	79.4	85.8	0.56
3				
Positive emotions+discrepancy+personal concerns	0.70	81.2	89	0.63
Analytical thinking score+other words+drives	0.69	79.4	86.5	0.55
Analytical thinking score+clout score+other words	0.68	76.4	88.8	0.45
Analytical thinking score+positive emotions+other words	0.67	79.4	86.1	0.56
Analytical thinking score+positive emotions+discrepancy	0.66	80	88.6	0.59
4				
Analytical thinking score+other words+cognitive processes+drives	0.72	78.2	87.5	0.52
Analytical thinking score+positive emotions+discrepancy+personal concerns	0.71	78.2	89.2	0.52
Positive emotions+causation+discrepancy+personal concerns	0.70	77.6	90	0.49
Analytical thinking score+clout score+first-person pronouns+positive emotions	0.70	80	88.8	0.59
Analytical thinking score+positive emotions+discrepancy+drives	0.70	78.2	87.6	0.52
5				
Analytical thinking score+positive emotions+causation+discrepancy+drives	0.73	80.6	87.9	0.60
Analytical thinking score+clout score+positive emotions+social processes+other words	0.72	80	88	0.59
Analytical thinking score+emotional tone score+discrepancy+other words+personal concerns	0.72	78.8	87.9	0.54
Analytical thinking score+positive emotions+discrepancy+other words+drives	0.72	80	87.5	0.59
Clout score+positive emotions+discrepancy+social processes+personal concerns	0.72	81.2	89.7	0.63
2[b] All predictors	0.45	34.7	32.1	0.30

[a]AUC: Area under the Receiver Operating Characteristic curve.

[b]The best models with 6-20 predictors (total of 75 items) have lower AUC, general accuracy, average accuracy, and F1 score, thus are omitted.

Further analysis was conducted on a subset of participants (43/105, 40.9%) who responded to the single-item mental health self-rating immediately after the chat session (49/165, 29.7%). On the basis of language use patterns, a binary classification procedure was able to distinguish those who rated their current mental health status as *poor* from those who rated their health as *fair* to *excellent* (2-5 out of 5) with high discrimination accuracy (AUC 0.95, general accuracy 85.7%, and average accuracy 88.7%), showing a better prediction of participant's mental health status shortly after a chat session.

All participants who had engaged in the text-based counseling service since August 2019 were contacted in July 2020 and asked to respond to the single-item rating of their mental health. This increased the number of responses from 49 to 165. Binary classification of this expanded data set yielded an acceptable accuracy rate (AUC 0.73); however, this was not as high as that for the participants who rated their mental health immediately after counseling had occurred. We interpret this as an encouraging indication that language use patterns are robust predictors of mental health status and that a larger data set of mental health ratings recorded immediately after counseling would likely yield excellent classification based on language use patterns. All analyses were rerun with the custom Australianized dictionary, but this did not improve the accuracy for any of the models.

Discussion

Principal Findings

This study aimed to determine whether language use patterns during the course of text-based counseling with a human therapist could be used to predict mental health status. Computational linguistic techniques were used to explore predictive relationships between language use patterns and the participants' underlying psychological presenting problem, which was recorded before the commencement of counseling, and the self-ratings of their current mental health status, which were recorded after counseling had been completed.

The two aims of the study were to investigate whether language use patterns could be used to (1) identify mental health presenting problems in the clients of *Virtual Psychologist* and (2) predict their self-reported mental health status. Our computational analysis was able to predict the top 3 presenting problems (anxiety, depression, and stress) with an accuracy of 80% (Table 11). The analysis was able to discriminate between those top 3 presenting problems with an accuracy ranging from 50% to 70% (Table 12), which was above the chance level. For the prediction of mental health status as determined by responses to the question, "How would you rate your mental health now?" the average accuracy of prediction was good, ranging from 80% to 90% (Table 13).

Language Use Patterns Can Be Used to Accurately Classify Presenting Problem and Future Mental Health Status

The findings suggest that language use patterns are useful indicators of mental health presenting problems and also predictive of future mental health status. We were able to use linguistic patterns to discriminate the top 3 presenting problems from the remaining pool of 17 presenting problems. This binary classification was able to separate participants with high accuracy. This finding is consistent with previous studies that have reported that depression has linguistic markers, such as increased use of first-person personal pronouns [40-44] and negative emotion words [51]. This study extends past findings by examining language use in a sample of participants with clinically significant presentation, who were receiving text-based counseling. In addition, the approach used here confirms the viability of using text-based counseling chat logs to enable computational linguistic analyses to determine the type of presenting problem.

The predictive analysis was successful in classifying the participants based on their self-ratings of mental health. Binary classification yielded high accuracy in identifying those participants who rated their mental health as *poor* following counseling versus those who rated it as *fair* to *excellent*. This finding provides compelling evidence that linguistic patterns are accurate and robust predictors of future mental health status. Our results support existing evidence that there are linguistic markers related to reductions in symptom severity and improved treatment outcomes [51-53]. To further improve accuracy, we recommend measuring participants' mental health in a standardized way to reduce variability introduced, for example, by differences in how long after the completion of counseling, the mental health status was measured; however, we acknowledge the difficulties in implementing these recommendations in a real-world clinical context.

As shown in Tables 11 and 13, several linguistic parameters emerged in both the best-performing models for predicting the presenting problems and self-rated mental well-being: first-person singular pronouns, emotional tones, and drives. These linguistic indicators seem to be more related to certain mental problems than other indicators found in existing studies conducted using other approaches [40-44,51-53]. Therefore, our results are consistent with those of previous studies. However, similar to many other computational approaches, one of the disadvantages of using discriminant analysis to find the best-performing prediction model with specific parameters is that, sometimes, it is difficult to interpret why certain parameters emerge as important predictors.

Unlike the study by Seabrook et al [55], the use of a customized dictionary did not improve the accuracy of our analyses. It is unclear why we did not observe similar improvements in depression identification as reported by those authors. There are several important differences between the 2 studies regarding the participant populations and methods of data collection. Seabrook et al [55] recruited participants from a younger age range, likely from urban areas; analyzed Facebook and Twitter status updates; and related these to the scores from a mood-tracking app that their participants downloaded and used. This differs markedly from the farmers recruited in this study, who were engaged in text-based psychological counseling. It is possible that the counseling context is less likely to elicit Australianism, such as slang and other colloquialisms, than posts on social media.

Differentiating Anxiety From Depression From Stress Is Difficult

Our observation that the models successfully differentiated the top 3 presenting problems from the rest, but were less accurate in differentiating among the top 3 presenting problems, requires explanation. This suggests that the top 3 presenting problems (anxiety, depression, and stress) share common features. This commonality may refer to both the linguistic patterns that individuals with these conditions use and the psychological symptoms that they exhibit. For instance, all 3 problems are likely to affect mood and motivation. Depression and anxiety are known to be highly comorbid conditions. Indeed, 45.7% of individuals with lifetime major depressive disorder also had a lifetime history of ≥1 anxiety disorders [67]. Depression and anxiety also commonly coexist [68]. Furthermore, stress is a response to pressures or threats, whereas anxiety may manifest as a reaction to the stress. Anxiety may not have a clear cause, and as a result, can last longer and be more difficult to treat, but at the time of presentation, both may be affecting the individual. Recall that our participants were recruited during the combined unprecedented events of the Australian bushfire season of 2019 to 2020 and the global COVID-19 pandemic. It seems entirely valid that it may not be possible to statistically differentiate anxiety, depression, and stress from one another because a sizable proportion of participants may have copresented with 2 or all 3 of these mental health problems simultaneously. Another explanation is that it is unknown whether these conditions can be differentiated using default LIWC categories. This is the first study to attempt to differentiate depression from anxiety from stress using linguistic patterns as computed by LIWC. Improved differentiation might be possible by refining the analysis to count the words with the strongest predictive power to separate one condition from the other (ie, going to a finer level of resolution than the coarse LIWC category), as has been demonstrated elsewhere [54]. These possibilities await to be tested in future studies.

Implications for Practice

The potential applications of an accurate, scalable approach to mental health are far-reaching, with implications for early screening and targeted interventions. Mental disorders are a leading cause of disability worldwide, with enormous economic consequences including lost productivity, employee absenteeism [69], and additional strain placed on carers [70] and health systems [71]. The economic costs owing to lost productivity and absenteeism, even in the case of mild depression, are estimated to be Aus $8 billion (US $5.68 billion) per annum [69]. Although natural language processing of electronic health records is increasingly being used to study mental illness [72], case notes written by therapists and clinicians do not capture the implicit nuances present in the language use patterns of their clients. Thus, they do not lend themselves to the types of predictive analyses described here. Being able to predict future mental health status would enable proactive and early identification of at-risk individuals and bolster harm minimization efforts. On the basis of the linguistic analyses and predictions introduced here, an automated application could be developed to run in the background after each text-based counseling session and output the possible presenting problems and mental well-being status. This could be useful for clinical psychologists to screen at-risk individuals at an early stage and provide subsequent intervention if needed. Thus, such an application would have the potential use of an assistive tool for clinical psychologists. The ultimate goal of such studies is to accurately predict which individuals are at risk of mental health problems (including suicide) so that mental health professionals can intervene and save that person's life. The present data offer the tantalizing possibility that text-based predictors of mental health status may enable large-scale automatic screening of mental illness and identification of at-risk individuals in the not-too-distant future.

Limitations and Future Directions

This study has several limitations. First, language use patterns were related to the participants' presenting problems and self-rated mental health status, but no neuropsychological assessments were administered. Given that our ultimate aim was to identify individuals at risk of clinically significant presentation, it would be advantageous to be able to relate language use patterns to standardized measures of psychological function.

Second, although three-fourth of the participants completed multiple sessions of counseling, the data set only permitted us to relate their language use patterns to presenting problems (recorded before the commencement of treatment) or self-reported mental health status (recorded after they had received counseling). Given that changes in language use have been observed during the course of treatment and these changes have been linked to treatment outcomes [73-75], it would be useful to also track such changes longitudinally when trying to determine a client's presenting problem. In addition, changes in language use could also be predictive of responsiveness to treatment and future mental health status. For example, increasing use of reflexive language and decreasing use of external language in therapeutic conversation have been associated with better therapeutic outcomes [73].

Third, although the size of the analyzed sample was considerable, the response rate for the self-rating of mental well-being was relatively low (165/773, 21.3% of the sessions). Future replications using similar approaches would benefit from larger data sets, which will increase statistical power and support the detection of significant associations between increased number of variables.

To address these limitations, future studies should include standardized neuropsychological assessments and pharmacological management histories. Ideally, these should be administered at multiple time points during the course of the study to measure changes in symptom severity and how they are reflected in the changes in language use.

Conclusions

This study suggests that language use patterns during the course of text-based counseling are robust predictors of mental health status in farmers living in rural and remote communities. Linguistic patterns can be used to accurately assign individuals into one of the top 3 presenting problem categories. They can also differentiate those top 3 presenting problems from the pool

of other presenting problems via binary classification. If replicated in other samples, computational linguistic analyses may be applied to big data approaches for mental health screening at the population level, providing insight into the linguistic patterns underlying the mental health needs of Australians and improving the speed and scale of identification of at-risk individuals. We were also able to accurately predict future mental health status (as measured by self-ratings) based on linguistic patterns. This technique can potentially provide a sensitive measure of future mental health status that may be used as an early indicator of being predisposed to mental health conditions such as depression, anxiety, and stress.

Text-based counseling serves an important treatment function and has the potential to span great distances to provide e-mental health services to areas where service capacity is lacking. Although text-based communication has limitations (slower than vocal exchanges, faceless, and impersonal), for some segments of the population, it is appealing because of those limits rather than in spite of them (low bandwidth and perceived as offering space and privacy). This study contributes to the understanding of the best approaches for using technology to promote mental well-being and identify individuals at risk of mental health problems.

Acknowledgments

This study was supported by funding from the New South Wales Government Centre for Work Health and Safety (tender SWNSW/654567/2019).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Australianized Linguistic Inquiry and Word Count dictionary.
[DOCX File , 18 KB - formative_v6i6e33036_app1.docx]

References

1. Vins H, Bell J, Saha S, Hess JJ. The mental health outcomes of drought: a systematic review and causal process diagram. Int J Environ Res Public Health 2015 Oct 22;12(10):13251-13275 [FREE Full text] [doi: 10.3390/ijerph121013251] [Medline: 26506367]
2. Mental health in rural and remote Australia. National Rural Health Alliance. 2017 Dec. URL: https://www.ruralhealth.org.au/sites/default/files/publications/nrha-mental-health-factsheet-dec-2017.pdf [accessed 2020-08-03]
3. Hazell T, Dalton H, Caton T, Perkins D. Rural suicide and its prevention: a CRRMH position paper. Centre for Rural and Remote Mental Health, University of Newcastle, Australia. 2017. URL: https://www.crrmh.com.au/content/uploads/RuralSuicidePreventionPaper_2017_WEB_FINAL.pdf [accessed 2020-08-03]
4. Andersen K, Hawgood J, Klieve H, Kõlves K, De Leo D. Suicide in selected occupations in Queensland: evidence from the State suicide register. Aust N Z J Psychiatry 2010 Mar;44(3):243-249. [doi: 10.3109/00048670903487142] [Medline: 20180726]
5. Kennedy AJ, Maple MJ, McKay K, Brumby SA. Suicide and accidental death in Australia's rural farming communities: a review of the literature. Rural Remote Health 2014;14(1):2517 [FREE Full text] [Medline: 24909931]
6. Milner A, Spittal MJ, Pirkis J, LaMontagne AD. Suicide by occupation: systematic review and meta-analysis. Br J Psychiatry 2013 Dec;203(6):409-416. [doi: 10.1192/bjp.bp.113.128405] [Medline: 24297788]
7. Austin EK, Handley T, Kiem AS, Rich JL, Lewin TJ, Askland HH, et al. Drought-related stress among farmers: findings from the Australian Rural Mental Health Study. Med J Aust 2018 Aug 20;209(4):159-165. [doi: 10.5694/mja17.01200] [Medline: 30041594]
8. Judd F, Jackson H, Komiti A, Murray G, Fraser C, Grieve A, et al. Help-seeking by rural residents for mental health problems: the importance of agrarian values. Aust N Z J Psychiatry 2006 Sep;40(9):769-776. [doi: 10.1080/j.1440-1614.2006.01882.x] [Medline: 16911752]
9. Perceval M, Ross V, Kõlves K, Reddy P, De Leo D. Social factors and Australian farmer suicide: a qualitative study. BMC Public Health 2018 Dec 12;18(1):1367 [FREE Full text] [doi: 10.1186/s12889-018-6287-7] [Medline: 30541526]
10. Benavides-Vaello S, Strode A, Sheeran BC. Using technology in the delivery of mental health and substance abuse treatment in rural communities: a review. J Behav Health Serv Res 2013 Jan;40(1):111-120. [doi: 10.1007/s11414-012-9299-6] [Medline: 23093443]
11. Hull TD, Mahan K. A study of asynchronous mobile-enabled SMS text psychotherapy. Telemed J E Health 2017 Mar;23(3):240-247. [doi: 10.1089/tmj.2016.0114] [Medline: 27797646]
12. DellaCrosse M, Mahan K, Hull TD. The effect of messaging therapy for depression and anxiety on employee productivity. J Technol Behav Sci 2018 Jun 26;4(1):1-5. [doi: 10.1007/s41347-018-0064-4]

13. Dwyer A, de Almeida Neto A, Estival D, Li W, Lam-Cassettari C, Antoniou M. Suitability of text-based communications for the delivery of psychological therapeutic services to rural and remote communities: scoping review. JMIR Ment Health 2021 Feb 24;8(2):e19478 [FREE Full text] [doi: 10.2196/19478] [Medline: 33625373]
14. Derks D, Fischer AH, Bos AE. The role of emotion in computer-mediated communication: a review. Comput Human Behav 2008 May;24(3):766-785. [doi: 10.1016/j.chb.2007.04.004]
15. Bargh JA, McKenna KY, Fitzsimons GM. Can you see the real me? Activation and expression of the "true self" on the Internet. J Soc Isssues 2002 Dec 17;58(1):33-48. [doi: 10.1111/1540-4560.00247]
16. Stubbings DR, Rees CS, Roberts LD. New avenues to facilitate engagement in psychotherapy: the use of videoconferencing and text–chat in a severe case of obsessive‐compulsive disorder. Aust Psychol 2015;50(4):265-270. [doi: 10.1111/ap.12111]
17. Fukkink RG, Hermanns JM. Children's experiences with chat support and telephone support. J Child Psychol Psychiatry 2009 Jun;50(6):759-766. [doi: 10.1111/j.1469-7610.2008.02024.x] [Medline: 19207634]
18. King R, Bambling M, Reid W, Thomas I. Telephone and online counselling for young people: a naturalistic comparison of session outcome, session impact and therapeutic alliance. Couns Psychother Res 2006 Sep;6(3):175-181. [doi: 10.1080/14733140600874084]
19. Rodda S, Lubman DI. Characteristics of gamblers using a national online counselling service for problem gambling. J Gambl Stud 2014 Jun;30(2):277-289. [doi: 10.1007/s10899-012-9352-7] [Medline: 23297169]
20. Beattie A, Shaw A, Kaur S, Kessler D. Primary-care patients' expectations and experiences of online cognitive behavioural therapy for depression: a qualitative study. Health Expect 2009 Mar;12(1):45-59 [FREE Full text] [doi: 10.1111/j.1369-7625.2008.00531.x] [Medline: 19250152]
21. Fukkink R, Hermanns J. Counseling children at a helpline: chatting or calling? J Community Psychol 2009 Nov;37(8):939-948. [doi: 10.1002/jcop.20340]
22. Kramer J, Conijn B, Oijevaar P, Riper H. Effectiveness of a Web-based solution-focused brief chat treatment for depressed adolescents and young adults: randomized controlled trial. J Med Internet Res 2014 May 29;16(5):e141 [FREE Full text] [doi: 10.2196/jmir.3261] [Medline: 24874006]
23. Cohen GE, Kerr BA. Computer-mediated counseling. Comput Human Serv 1999 Apr 22;15(4):13-26. [doi: 10.1300/J407v15n04_02]
24. Andersson G. Internet-delivered psychological treatments. Annu Rev Clin Psychol 2016;12:157-179. [doi: 10.1146/annurev-clinpsy-021815-093006] [Medline: 26652054]
25. Andersson G, Carlbring P, Titov N, Lindefors N. Internet interventions for adults with anxiety and mood disorders: a narrative umbrella review of recent meta-analyses. Can J Psychiatry 2019 Jul;64(7):465-470 [FREE Full text] [doi: 10.1177/0706743719839381] [Medline: 31096757]
26. Andersson G, Rozental A, Shafran R, Carlbring P. Long-term effects of Internet-supported cognitive behaviour therapy. Expert Rev Neurother 2018 Jan;18(1):21-28. [doi: 10.1080/14737175.2018.1400381] [Medline: 29094622]
27. Crutzen R, Bosma H, Havas J, Feron F. What can we learn from a failed trial: insight into non-participation in a chat-based intervention trial for adolescents with psychosocial problems. BMC Res Notes 2014 Nov 20;7:824 [FREE Full text] [doi: 10.1186/1756-0500-7-824] [Medline: 25409911]
28. Kordy H, Wolf M, Aulich K, Bürgy M, Hegerl U, Hüsing J, et al. Internet-delivered disease management for recurrent depression: a multicenter randomized controlled trial. Psychother Psychosom 2016;85(2):91-98. [doi: 10.1159/000441951] [Medline: 26808817]
29. Babu S. Predicting depression from social media updates. University of California, Irvine. 2018 Jun. URL: http://asterix.ics.uci.edu/thesis/Sharon_Babu_Honors_thesis_2018.pdf [accessed 2020-08-03]
30. De Choudhury M, Gamon M, Counts S, Horvitz E. Predicting depression via social media. In: Proceedings of the 7th International AAAI Conference on Web and Social Media. 2013 Presented at: AAAI '13; July 8-11, 2013; Boston, MA, USA p. 8-11.
31. Guntuku SC, Yaden DB, Kern ML, Ungar LH, Eichstaedt JC. Detecting depression and mental illness on social media: an integrative review. Curr Opin Behav Sci 2017 Dec;18:43-49. [doi: 10.1016/j.cobeha.2017.07.005]
32. Ruiz V, Shi L, Quan W, Ryan N, Biernesser C, Brent D, et al. CLPsych2019 shared task: predicting suicide risk level from reddit posts on multiple forums. In: Proceedings of the 6th Workshop on Computational Linguistics and Clinical Psychology. 2019 Presented at: CLPsych '19; June 6, 2019; Minneapolis, MN, USA p. 162-166. [doi: 10.18653/v1/w19-3020]
33. Schwartz HA, Eichstaedt J, Kern ML, Park G, Sap M, Stillwell D, et al. Towards assessing changes in degree of depression through Facebook. In: Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality. 2014 Presented at: CLPsych '14; June 27, 2014; Baltimore, MD, USA p. 118-125. [doi: 10.3115/v1/w14-3214]
34. Al-Mosaiwi M, Johnstone T. In an absolute state: elevated use of absolutist words is a marker specific to anxiety, depression, and suicidal ideation. Clin Psychol Sci 2018 Jul;6(4):529-542 [FREE Full text] [doi: 10.1177/2167702617747074] [Medline: 30886766]
35. Davcheva E. Text mining mental health forums - learning from user experiences. In: Proceedings of the 26th European Conference on Information Systems: Beyond Digitization - Facets of Socio-Technical Change. 2018 Presented at: ECIS '18; June 23-28, 2018; Portsmouth, UK p. 91-101.

36. Fekete S. The Internet - a new source of data on suicide, depression and anxiety: a preliminary study. Arch Suicide Res 2002 Nov;6(4):351-361. [doi: 10.1080/13811110214533]
37. Park A, Conway M, Chen AT. Examining thematic similarity, difference, and membership in three online mental health communities from Reddit: a text mining and visualization approach. Comput Human Behav 2018 Jan;78:98-112 [FREE Full text] [doi: 10.1016/j.chb.2017.09.001] [Medline: 29456286]
38. Pennebaker JW, Boyd RL, Jordan K, Blackburn K. The development and psychometric properties of LIWC2015. The University of Texas at Austin. 2015. URL: https://repositories.lib.utexas.edu/bitstream/handle/2152/31333/LIWC2015_LanguageManual.pdf?sequence=3&isAllowed=y [accessed 2019-06-27]
39. Tausczik YR, Pennebaker JW. The psychological meaning of words: LIWC and computerized text analysis methods. J Lang Soc Psychol 2009 Dec 08;29(1):24-54. [doi: 10.1177/0261927X09351676]
40. Fast LA, Funder DC. Gender differences in the correlates of self-referent word use: authority, entitlement, and depressive symptoms. J Pers 2010 Feb;78(1):313-338. [doi: 10.1111/j.1467-6494.2009.00617.x] [Medline: 20433621]
41. Molendijk ML, Bamelis L, van Emmerik AA, Arntz A, Haringsma R, Spinhoven P. Word use of outpatients with a personality disorder and concurrent or previous major depressive disorder. Behav Res Ther 2010 Jan;48(1):44-51. [doi: 10.1016/j.brat.2009.09.007] [Medline: 19819423]
42. Pyszczynski T, Greenberg J. Toward an integration of cognitive and motivational perspectives on social inference: a biased hypothesis-testing model. In: Berkowitz L, editor. Advances in Experimental Social Psychology. Cambridge, MA, USA: Academic Press; 1987:297-340.
43. Rude S, Gortner EM, Pennebaker J. Language use of depressed and depression-vulnerable college students. Cogn Emot 2004 Dec;18(8):1121-1133. [doi: 10.1080/02699930441000030]
44. Weintraub W. Verbal Behavior: Adaptation and Psychopathology. Berlin, Germany: Springer; 1981.
45. Brockmeyer T, Zimmermann J, Kulessa D, Hautzinger M, Bents HC, Friederich H, et al. Me, myself, and I: self-referent word use as an indicator of self-focused attention in relation to depression and anxiety. Front Psychol 2015 Oct 9;6:1564 [FREE Full text] [doi: 10.3389/fpsyg.2015.01564] [Medline: 26500601]
46. Zimmermann J, Brockmeyer T, Hunn M, Schauenburg H, Wolf M. First-person pronoun use in spoken language as a predictor of future depressive symptoms: preliminary evidence from a clinical sample of depressed patients. Clin Psychol Psychother 2017 Mar;24(2):384-391. [doi: 10.1002/cpp.2006] [Medline: 26818665]
47. Lyons M, Aksayli ND, Brewer G. Mental distress and language use: linguistic analysis of discussion forum posts. Comput Human Behav 2018 Oct;87:207-211. [doi: 10.1016/j.chb.2018.05.035]
48. Tackman AM, Sbarra DA, Carey AL, Donnellan MB, Horn AB, Holtzman NS, et al. Depression, negative emotionality, and self-referential language: a multi-lab, multi-measure, and multi-language-task research synthesis. J Pers Soc Psychol 2019 May;116(5):817-834. [doi: 10.1037/pspp0000187] [Medline: 29504797]
49. Stirman SW, Pennebaker JW. Word use in the poetry of suicidal and nonsuicidal poets. Psychosom Med 2001;63(4):517-522. [doi: 10.1097/00006842-200107000-00001] [Medline: 11485104]
50. Dirkse D, Hadjistavropoulos HD, Hesser H, Barak A. Linguistic analysis of communication in therapist-assisted Internet-delivered cognitive behavior therapy for generalized anxiety disorder. Cogn Behav Ther 2015;44(1):21-32. [doi: 10.1080/16506073.2014.952773] [Medline: 25244051]
51. Arntz A, Hawke LD, Bamelis L, Spinhoven P, Molendijk ML. Changes in natural language use as an indicator of psychotherapeutic change in personality disorders. Behav Res Ther 2012 Mar;50(3):191-202. [doi: 10.1016/j.brat.2011.12.007] [Medline: 22317755]
52. Van der Zanden R, Curie K, Van Londen M, Kramer J, Steen G, Cuijpers P. Web-based depression treatment: associations of clients' word use with adherence and outcome. J Affect Disord 2014 May;160:10-13 [FREE Full text] [doi: 10.1016/j.jad.2014.01.005] [Medline: 24709016]
53. Huston J, Meier S, Faith M, Reynolds A. Exploratory study of automated linguistic analysis for progress monitoring and outcome assessment. Couns Psychother Res 2019 Apr 16;19(3):321-328. [doi: 10.1002/capr.12219]
54. Owen JE, Klapow JC, Roth DL, Shuster Jr JL, Bellis J, Meredith R, et al. Randomized pilot of a self-guided Internet coping group for women with early-stage breast cancer. Ann Behav Med 2005 Aug;30(1):54-64. [doi: 10.1207/s15324796abm3001_7] [Medline: 16097906]
55. Seabrook EM, Kern ML, Fulcher BD, Rickard NS. Predicting depression from language-based emotion dynamics: longitudinal analysis of Facebook and Twitter status updates. J Med Internet Res 2018 May 08;20(5):e168 [FREE Full text] [doi: 10.2196/jmir.9267] [Medline: 29739736]
56. Calvo RA, Milne DN, Hussain MS, Christensen H. Natural language processing in mental health applications using non-clinical texts. Nat Lang Eng 2017 Jan 30;23(5):649-685. [doi: 10.1017/S1351324916000383]
57. Australia. Virtual Psychologist. 2020. URL: https://www.virtualpsychologist.com.au [accessed 2020-09-25]
58. Althoff T, Clark K, Leskovec J. Large-scale analysis of counseling conversations: an application of natural language processing to mental health. Trans Assoc Comput Linguist 2016;4:463-476 [FREE Full text] [Medline: 28344978]
59. First confirmed case of novel coronavirus in Australia. Australian Government. 2020. URL: https://www.health.gov.au/ministers/the-hon-greg-hunt-mp/media/first-confirmed-case-of-novel-coronavirus-in-australia [accessed 2020-08-03]

60. Beautrais AL. Farm suicides in New Zealand, 2007-2015: a review of coroners' records. Aust N Z J Psychiatry 2018 Jan;52(1):78-86. [doi: 10.1177/0004867417704058] [Medline: 28410567]
61. Pennebaker JW, Chung CK, Frazee J, Lavergne GM, Beaver DI. When small words foretell academic success: the case of college admissions essays. PLoS One 2014 Dec 31;9(12):e115844 [FREE Full text] [doi: 10.1371/journal.pone.0115844] [Medline: 25551217]
62. Kacewicz E, Pennebaker JW, Davis M, Jeon M, Graesser AC. Pronoun use reflects standings in social hierarchies. J Lang Soc Psychol 2013 Sep 19;33(2):125-143. [doi: 10.1177/0261927x13502654]
63. Newman ML, Pennebaker JW, Berry DS, Richards JM. Lying words: predicting deception from linguistic styles. Pers Soc Psychol Bull 2003 May 02;29(5):665-675. [doi: 10.1177/0146167203029005010] [Medline: 15272998]
64. Pennebaker JW. The Secret Life of Pronouns: What Our Words Say About Us. New York, NY, USA: Bloomsbury Press/Bloomsbury Publishing; 2011.
65. Cohn MA, Mehl MR, Pennebaker JW. Linguistic markers of psychological change surrounding September 11, 2001. Psychol Sci 2004 Oct;15(10):687-693. [doi: 10.1111/j.0956-7976.2004.00741.x] [Medline: 15447640]
66. Rice ME, Harris GT. Comparing effect sizes in follow-up studies: ROC Area, Cohen's d, and r. Law Hum Behav 2005 Oct;29(5):615-620. [doi: 10.1007/s10979-005-6832-7] [Medline: 16254746]
67. Kessler RC, Sampson NA, Berglund P, Gruber MJ, Al-Hamzawi A, Andrade L, et al. Anxious and non-anxious major depressive disorder in the World Health Organization World Mental Health Surveys. Epidemiol Psychiatr Sci 2015 Jun;24(3):210-226 [FREE Full text] [doi: 10.1017/S2045796015000189] [Medline: 25720357]
68. Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshleman S, et al. Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Results from the National Comorbidity Survey. Arch Gen Psychiatry 1994 Jan;51(1):8-19. [doi: 10.1001/archpsyc.1994.03950010008002] [Medline: 8279933]
69. McTernan WP, Dollard MF, LaMontagne AD. Depression in the workplace: an economic cost analysis of depression-related productivity loss attributable to job strain and bullying. Work Stress 2013 Nov 07;27(4):321-338. [doi: 10.1080/02678373.2013.846948]
70. Briggs H, Fisher D. Warning: Caring is a Health Hazard : Results of the 1999 National Survey of Carer Health and Wellbeing. Turner, Australia: Carers Association of Australia; 2000.
71. Lee YC, Chatterton ML, Magnus A, Mohebbi M, Le LK, Mihalopoulos C. Cost of high prevalence mental disorders: findings from the 2007 Australian National Survey of Mental Health and Wellbeing. Aust N Z J Psychiatry 2017 Dec;51(12):1198-1211. [doi: 10.1177/0004867417710730] [Medline: 28565923]
72. Edgcomb JB, Zima B. Machine learning, natural language processing, and the electronic health record: innovations in mental health services research. Psychiatr Serv 2019 Apr 01;70(4):346-349. [doi: 10.1176/appi.ps.201800401] [Medline: 30784377]
73. Banham JA, Schweitzer RD. Comparative exploration of narrative processes for better and poorer outcomes for depression. Couns Psychother Res 2015;15(3):228-238. [doi: 10.1002/capr.12032]
74. Bento T, Ribeiro AP, Salgado J, Mendes I, Gonçalves MM. The narrative model of therapeutic change: an exploratory study tracking innovative moments and protonarratives using state space grids. J Constr Psychol 2013 Dec 13;27(1):41-58. [doi: 10.1080/10720537.2014.850373]
75. Reyes L, Aristegui R, Krause M, Strasser K, Tomicic A, Valdes N, et al. Language and therapeutic change: a speech acts analysis. Psychother Res 2008 May;18(3):355-362. [doi: 10.1080/10503300701576360] [Medline: 18815987]

Abbreviations

AUC: Area under the Receiver Operating Characteristic curve
LIWC: Linguistic Inquiry and Word Count

Edited by A Mavragani; submitted 19.08.21; peer-reviewed by T Hull, DMSH Mukta, J Feldhege; comments to author 04.10.21; revised version received 26.11.21; accepted 21.04.22; published 17.06.22.

Please cite as:
Antoniou M, Estival D, Lam-Cassettari C, Li W, Dwyer A, Neto ADA
Predicting Mental Health Status in Remote and Rural Farming Communities: Computational Analysis of Text-Based Counseling
JMIR Form Res 2022;6(6):e33036
URL: https://formative.jmir.org/2022/6/e33036/
doi:10.2196/33036
PMID:

Original Paper

The Drivers of Acceptance of Artificial Intelligence–Powered Care Pathways Among Medical Professionals: Web-Based Survey Study

Lisa Cornelissen[1], MSc; Claudia Egher[1,2], PhD; Vincent van Beek[3,4], MD; Latoya Williamson[3], MSc; Daniel Hommes[3,4], MD, Prof Dr

[1]Faculty of Science, Athena Institute, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

[2]Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, Netherlands

[3]DEARhealth, Amsterdam, Netherlands

[4]Department of Gastroenterology & Hepatology, Leiden University Medical Center, Leiden, Netherlands

Corresponding Author:
Lisa Cornelissen, MSc
Faculty of Science
Athena Institute
Vrije Universiteit Amsterdam
Boelelaan 1105
Amsterdam, 1081HV
Netherlands
Phone: 31 655320046
Email: lisa.cornelissen@dearhealth.com

Abstract

Background: The emergence of Artificial Intelligence (AI) has been proven beneficial in several health care areas. Nevertheless, the uptake of AI in health care delivery remains poor. Despite the fact that the acceptance of AI-based technologies among medical professionals is a key barrier to their implementation, knowledge about what informs such attitudes is scarce.

Objective: The aim of this study was to identify and examine factors that influence the acceptability of AI-based technologies among medical professionals.

Methods: A survey was developed based on the Unified Theory of Acceptance and Use of Technology model, which was extended by adding the predictor variables perceived trust, anxiety and innovativeness, and the moderator profession. The web-based survey was completed by 67 medical professionals in the Netherlands. The data were analyzed by performing a multiple linear regression analysis followed by a moderating analysis using the Hayes PROCESS macro (SPSS; version 26.0, IBM Corp).

Results: Multiple linear regression showed that the model explained 75.4% of the variance in the acceptance of AI-powered care pathways (adjusted R^2=0.754; $F_{9,0}$=22.548; P<.001). The variables medical performance expectancy (β=.465; P<.001), effort expectancy (β=–.215; P=.005), perceived trust (β=.221; P=.007), nonmedical performance expectancy (β=.172; P=.08), facilitating conditions (β=–.160; P=.005), and professional identity (β=.156; P=.06) were identified as significant predictors of acceptance. Social influence of patients (β=.042; P=.63), anxiety (β=.021; P=.84), and innovativeness (β=.078; P=.30) were not identified as significant predictors. A moderating effect by gender was found between the relationship of facilitating conditions and acceptance (β=–.406; P=.09).

Conclusions: Medical performance expectancy was the most significant predictor of AI-powered care pathway acceptance among medical professionals. Nonmedical performance expectancy, effort expectancy, perceived trust, and professional identity were also found to significantly influence the acceptance of AI-powered care pathways. These factors should be addressed for successful implementation of AI-powered care pathways in health care delivery. The study was limited to medical professionals in the Netherlands, where uptake of AI technologies is still in an early stage. Follow-up multinational studies should further explore the predictors of acceptance of AI-powered care pathways over time, in different geographies, and with bigger samples.

(JMIR Form Res 2022;6(6):e33368) doi:10.2196/33368

KEYWORDS

technology acceptance; artificial intelligence; health care providers; machine learning; technology adoption; health innovation; user adoption

Introduction

Health care systems are currently burdened owing to an aging population, increasing life expectancy, the development of expensive therapies, an inefficient design, and a growing demand for a good quality of care [1]. This has resulted in rising health care expenditure and threatened the accessibility of care [1]. Artificial Intelligence (AI), broadly defined as the capability of a machine to imitate intelligent human behavior [2], has the potential to help improve many of these challenges. Through the development of sophisticated algorithms, AI can assist in the diagnosis, monitoring, and treatment of patients, it can help streamline services and render administrative tasks more efficient [3]. Even though AI has already been proven beneficial in several health areas, such as clinical decision support, patient monitoring, health interventions, and health care administration [4-6], its impact on health care delivery has thus far remained limited [7].

Several barriers for entry have been identified, which explain the underuse of AI in health care delivery, including regulatory constraints, ethical considerations, lack of transparency, and the lack of facilitating conditions [8-10]. In addition, a crucial barrier for the implementation of AI-based technologies is the lack of adoption among medical professionals [9]. Moreover, individuals' acceptance and utilization of technologies are proposed to be the most important factors for health technology adoption [11]. Currently, it is poorly understood what the reasons are for medical professionals (not) adopting AI technologies. Recently, some studies have been conducted to research the perspectives of the end users in the implementation of AI-based technologies, but more insight is essential [12].

The lack of understanding as to what informs the resistance among medical professionals in regard to the adoption of AI-based technologies can have important negative consequences, as it can limit and delay substantial improvements in health care delivery and result in wasted research and high design costs. Therefore, this study investigated medical professionals' perspectives on the adoption of AI-powered care pathways by identifying which factors influence, and to what degree, the acceptability of AI-based technologies among these stakeholders.

This study focused on AI-powered care pathway technology. This technology enables the management of chronic diseases on a digital platform. All stakeholders (including medical professionals, patients, and caregivers), involved medical activities, and associated support programs are included in this platform. It enables medical professionals to constantly monitor their patient population's disease activity and mental well-being through a patient app. The care pathways were designed to offer the right care at the right time and are continuously risk-adjusted using AI. This risk adjustment is created through several steps. The first step entails updating the patient's data into the system. In the second step, the data are classified in the patient's risk profile. In the third step, the risk profile–learning models and algorithms (based on AI) update the care pathway upon the most recent profile of the patient. From the update, a new recommendation is formulated (or not, if no alteration is necessary). Lastly, the medical professional can accept or reject the received recommendation based on various considerations and in dialogue with the patient.

Methods

Recruitment

The target population consisted of medical professionals who were employed at a Dutch hospital or other hospital staff who worked on improvement of the quality of care. Participants were mainly invited to participate through the physicians' network (email, LinkedIn, virtual, and in-person meetings). A web-based survey was created using Qualtrics [13]. The data were gathered between the April 20 and June 1, 2021. The survey took approximately 7 minutes to complete.

Model

To determine factors that influence the acceptability of AI-powered care pathways among medical professionals, the validated Unified Theory of Acceptance and Use of Technology (UTAUT) model was extended and subsequently applied [14]. According to the UTAUT model, the acceptance of new technology can be measured by the behavioral intention (BI) to use a certain technology. The following predictor variables from UTAUT were included for the analysis: performance expectancy (PE; divided into medical and nonmedical), effort expectancy (EE), social influence (SI; divided in to social influence patients and medical), and facilitating conditions (FC) [14]. The original construct performance expectancy was divided into medical and nonmedical since Shaw et al [15] stated that AI-based technologies have different relevant tasks (clinical, epidemiological, and operational), and uncovering the value proposition between these tasks is an essential consideration for successful adoption. To uncover the value proposition for AI-powered care pathways, the construct of performance expectancy medical (*clinical* in article of Shaw et al [15]) and performance expectancy nonmedical (*operational* in article of Shaw et al [15]). The construct of social influence was divided since different studies highlighted that social influence is often studied from one influential group while neglecting influence of other groups [3,4,16,17]. Eckhardt et al [16] proposed to derive relevant influential groups and treat their different impacts with due respect. In light of this study, two main influential groups were identified, namely medical *professionals* and *patients*, resulting in the following constructs: social influence medical experts (SIME) and social influence patients (SIPA). The model was enriched with several variables that relevant scientific literature from multiple disciplines identified as playing a role in shaping the technology acceptance of AI-based technologies. These variables were the following: perceived trust (PT) [18], anxiety (AN) [19], professional

Identity (PI) [20], and innovativeness (IN) [21] (Table 1). Furthermore, three moderators from the UTAUT model were included, namely age, gender, and experience [14]. In the original model, experience is defined as experience with the used technology. In the light of this study, this moderator was not applicable since AI-powered care pathways are in a premature stage of implementation. Therefore, the definition was changed to "Years of experience in the medical field" [22]. An additional moderator, *profession*, was added since different professions have different responsibilities and tasks, which is hypothesized to influence the relationships between the predictor variables and the acceptance. A schematic overview of the model is shown in Figure 1.

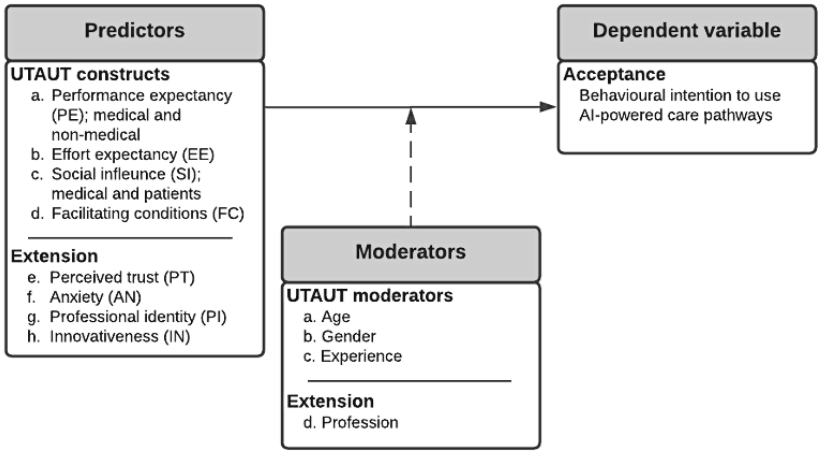
Table 1. Definitions of the predictor variables for the behavioral intention to use artificial intelligence (AI)-powered care pathways.

Construct	Operational definition
Medical performance expectancy[a]	Degree to which an individual believes that using AI-powered care pathways will help him or her to attain gains in terms of the provided quality of care [14,15]
Nonmedical performance expectancy[a]	Degree to which an individual believes that using AI-powered care pathways will help him or her to attain gains in productivity, efficiency, and communication [14,15]
Effort expectancy	Degree of ease associated with the use of the system [14]
Social influence patients[b]	Degree to which an individual perceives that patients believe that he or she should use the new system [14,16,17]
Social influence medical[b]	Degree to which an individual perceives that other medical organizations or colleagues believe that he or she should use the new system [14,16,17]
Facilitating conditions	Degree to which an individual believes that an organizational and technical infrastructure exists to support the use of the system [14]
Perceived trust	Users' specific trust that AI-powered care pathways have the ability, integrity, and benevolence in providing their service [18]
Anxiety	The fear (eg, sadness, perception, and stress caused by stress-creating situations) experienced by an individual during their interaction with AI-powered care pathways [19]
Professional identity	The attitudes, values, knowledge, beliefs, and skills that are shared with others within a professional role being undertaken by the individual [23]
Innovativeness	Degree to which an individual is relatively earlier in adopting an innovation than other members of his (social) system [21]

[a]The original determinant in the Unified Theory of Acceptance and Use of Technology (UTAUT) model of performance expectancy was divided in two separate variables since performance expectancy for AI-powered care pathways can be viewed from a medical and nonmedical perspective.

[b]The original determinant in the UTAUT model of social influence was divided in two separate variables since it is hypothesized that patients and medical organizations or colleagues have different influences.

Figure 1. Overview of the conceptual model used in this study. The predictor variables (performance expectancy, effort expectancy, social influence, facilitating conditions, perceived trust, anxiety, professional identity, and innovativeness) are hypothesized to influence the variance the acceptance of AI-powered care pathways. The moderators (age, gender, experience, and profession) are hypothesized to influence the relationship between the predicator variables and the dependent variable. AI: artificial intelligence, UTAUT: Unified Theory of Acceptance and Use of Technology.



Survey

The survey contained questions about demographics including age, gender, experience, and profession. Then, the survey participants were invited to rate statements concerning the constructs that needed to be ranked using a 5-point Likert scale (1=totally disagree, 5=totally agree). The survey items were formulated by adopting statements from prior research and by

developing new statements within the research group (Multimedia Appendix 1 for statements). Before collecting the data, the survey was tested extensively through a pilot study with graduate students, individuals who were not familiar with AI-powered care pathways, and DEARhealth. staff. This pilot tested for confusing formulations, lay-out problems, the approximate time to complete the survey, and the technical resources needed. Where needed, adjustments were subsequently made.

Statistical Analysis

Measurement Model Testing

To assess the reliability of the measurements, the internal consistency was tested using Cronbach α values. The commonly used rule of thumb for Cronbach α was used where a value is acceptable above .6, questionable between .5 and .6, and unacceptable below .5 [24]. Methods to try and ensure reliability such as item removal were performed when problems of reliability arose. Furthermore, Pearson correlations between predictor variables were tested to rule out any internal relations, a rule of thumb of >0.7 was used.

Relationship Testing

The data were analyzed using SPSS (version 26; IBM Corp) including the extension PROCESS macro developed by [25]. To exclude responses with missing data, a data clean was conducted. A descriptive analysis to get an insight into the respondents' demographic characteristics was conducted. Then, a multiple linear regression analysis was performed to test the contribution of each predictor variable on the variance of BI. Before conducting the multiple linear regression analysis, the assumptions were checked to rule out violations. The checked violations were linearity, multivariate outliers, heteroscedasticity, and multicollinearity. A *P* value of <.01 was considered significant (Multimedia Appendix 2). Lastly, moderation analysis was conducted using PROCESS model 1. The moderation analysis used hierarchical multiple regression with an interaction term.

Ethical Considerations

No additional ethical approval was needed according to the online check performed using the web-based BETCHIE test of the Beta faculty of Vrije Universiteit Amsterdam, which indicated that the target group was not considered a vulnerable group in this research. The privacy of the respondents was ensured by anonymizing the survey in Qualtrics. The researchers could not track the source of the survey, and no private information was collected. Participating in this research was voluntary.

Results

Participants

In total, 111 health care professionals started the survey. After excluding respondents with missing answers (n=41) or monotone answers (n=3), 67 remained. Of the 67 participants, 41 (61.2%) identified as female and 26 (38.8%) as male. The age distribution in this research was the following: <35 years (20/67, 29.9%), 35-55 years (n=27, 40.3%), and >55 years (n=20, 29.9%). For the different medical professions, an overrepresentation of physicians (n=28, 41.8%) was identified as compared to the number of nurses (n=14, 20.9%) and nurse specialists (n=2, 3.0%) (Table 2).

Table 2. Participant demographics (N=67).

Characteristics	Participants, n	Participants, %
Gender		
Male	26	38.8
Female	41	61.2
Age (years)		
18-24	5	7.5
25-34	15	22.4
35-44	16	23.9
45-54	11	16.4
55-64	17	25.4
≥65	3	4.5
Experience in the medical field		
≤2	5	7.5
3-5	10	14.9
6-10	7	10.4
11-20	19	28.4
21-30	14	20.9
≥31	12	17.9
Profession		
Physician	28	41.8
Nurse specialist	2	3.0
Nurse	14	20.9
Management	11	16.4
Consultant	11	16.4
Other function in hospital	13	19.4

Outcomes

Measurement Testing Findings

A Cronbach α score was calculated for each construct of the model to validate the internal consistency of the measurement statements within the variable (Table 3). The variable of SIME showed a Cronbach α below .50 and was removed from the analysis. The Pearson correlation coefficients between any of the predictor variables did not exceed 0.7, indicating an acceptable correlation between the predictors (Multimedia Appendix 3).

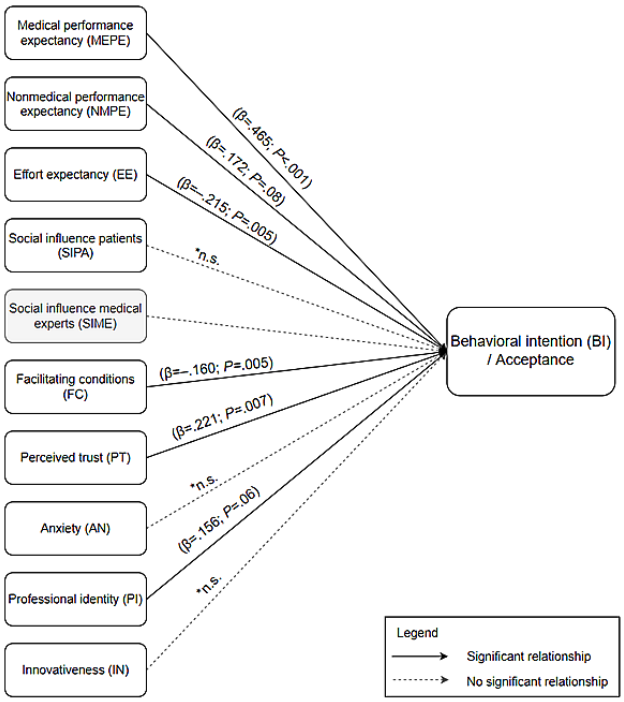
Table 3. Internal reliability of the constructs based on the 3 statements using Cronbach α values. Social influence medical experts (SIME) and facilitating conditions (FCs) showed unacceptable internal consistency (Cronbach α<.5). Item removal resulted in a better Cronbach α for facilitating conditions.

Variable	Cronbach α
Innovativeness	.706
Anxiety	.701
FC→FC1+FC2 (item removal of FC3)	.455 → .512
Nonmedical performance expectancy	.662
Social influence patients	.667
Medical performance expectancy	.638
Social influence medical experts	.244
Professional identity	.748
Perceived trust	.717
Effect expectancy	.816
Behavioral intention	.916

Regression Outcomes

The results of multiple linear regression analysis showed significant relationships between the predictor variables and the acceptance of AI-powered care pathways. Overall, the results show that 75.4% of the variance in the acceptance can be explained by the independent variables of the model (adjusted R^2=0.754; $F_{9,0}$=22.548; P<.001). From the data, it can be concluded that the model is highly significant (P<.001). The analysis indicated that the variables medical performance expectancy (MEPE; β=.465; P<.001), nonmedical performance expectancy (NMPE; β=.172; P=.08), PT (β=.221; P=.007), and PI (β=.156; P=.06) had a significant positive effect on the acceptance of AI-powered care pathways (Figure 2 and Multimedia Appendix 4). Both EE (β=–.215; P=.005) and FC (β=–.160; P=.005) were found to have a negative impact on acceptance. From the magnitude of the β statistics, MEPE had the biggest impact on variance followed by PT, EE, NMPE, FC, and PI. Some variables did not show a significant result, including SIPA (β=.042; P=.63), AN (β=.021; P=.84), and IN (β=.078; P=.30).

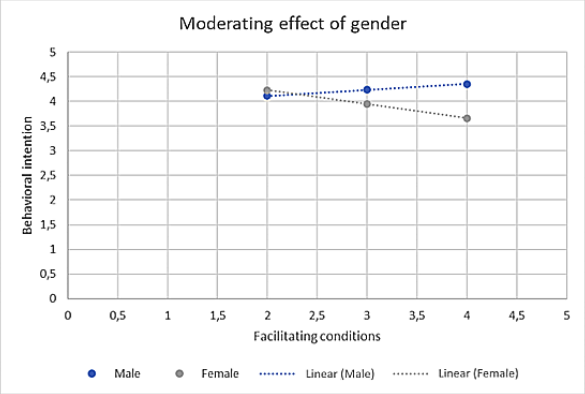
Figure 2. Overview of the individual relationships of the predictor variables and the acceptance of artificial intelligence–powered care pathways. Medical performance expectancy (MEPE), nonmedical performance expectancy, effort expectancy, facilitating condition, perceived trust, and professional identity showed a significantly influential relationship on acceptance where MEPE had the largest impact. Social influence patient, anxiety, and innovativeness did not show a significant relationship on the variance in acceptance. The predicator variable social influence medical was excluded from the analysis since it showed a poor internal consistency. n.s.: not significant.



Moderating Outcomes

The moderating effects of gender, age, experience, and profession were each tested for the relationships between the individual predictor variables and the acceptance to use AI-powered care pathways (Multimedia Appendix 5). Gender had a significant moderating effect on the relationship between facilitating conditions and the acceptance to use AI-powered care pathways (β=–.406; *P*=.09), indicating that being male had a positive moderating effect and female had a negative moderating effect (Figure 3). No other significant moderators were identified.

Figure 3. Moderating effect of gender. Being a male had a positive moderating effect whereas being a female had a negative impact.



Discussion

Principal Findings

This study investigated the technology acceptance of AI-powered care pathways among medical professionals. The model explained 75.4% of the variance in acceptance of the medical professionals. The predictor variables MEPE, NMPE, EE, FC, PT, and PI were found to significantly influence the acceptance of AI-powered care pathways. SIPA, SIME, IN, and AN were not found to significantly influence the behavioral intention. One moderating relationship was found; gender moderates the relationship between FC and acceptance, with identifying as male increasing the likelihood of accepting AI-powered care pathways.

Comparison With Prior Work

The predictor MEPE was found to have the highest impact on the acceptance of AI-powered care pathways. Several studies on the acceptance of health technology also identified performance expectancy as the main predictor [12,14,26-29]. The Predictor PEME was also found to be most important goal of physicians in the qualitative study of Lai et al [12]—they concluded that providing the best care for the patients was the main goal of physicians, and if AI-based technologies could enhance that, they were not opposed to change and the use of AI-based technologies. These findings confirm that medical professionals are more willing to use AI-powered care pathways when they see the benefits and added value considering the quality of care. Interestingly, when comparing the magnitude of the β values, MEPE was found to have more than double the influential strength compared to NMPE, implying that the perceived added value in terms of work efficiency, productivity, and communication has a lower impact on the acceptance than the perceived added value for the quality of care. This finding is in line with that of Shaw et al [15], where they stated that it is important to look at the value proposition between different added values a technology can bring. This finding suggests that the medical professionals in this study focus on the clinical relevance of AI-powered care pathways and that they would be more interested in the integrated care approach they would facilitate. Another possible explanation for this considerable difference in magnitude could be that medical professionals are less aware about the added value that AI-powered care pathways have regarding efficiency, productivity, and communication as most information about AI in health care tends to focus on the impact it can have on clinical aspects. This might have resulted in NMPE being less influential than MEPE in the variance of the acceptance.

A positive impact of PI was found, implying that if medical professionals perceive AI-powered care pathways as a positive stimulus to their career growth, professional status, and financial situation, they would be more willing to use the technology, and vice versa. This finding confirms the results of Jussupow et al [20], who proposed that for the successful implementations of information technology, especially AI in health care, it is crucial to identify and address professional identity threats. Currently, none of the popular technology acceptance models (UTAUT, Technology Acceptance Model, Diffusion of Innovation, and Technology Readiness Phases) include a determinant considering the influence of PI. This study highlights the importance of involving PI when studying the acceptance of target groups with a strong PI or social status and when dealing with AI-based technologies.

Social influence was not found to be an influential factor (both SIME and SIPA), which is contrary to several other studies that highlighted its importance for the adoption of new technologies [14,30]. The lack of impact of social influence in this study could be explained by the premature-stage AI-powered care pathways that the technology is in, as key opinion leaders are still absent and insufficient successful examples are present [31]. Future research should confirm if this is indeed the case. Furthermore, COVID-19 could have lowered the prioritization of AI-based technologies since the pandemic pressured the medical professionals, and no additional time was available to focus on AI-based technologies. This may have shaped their

priorities and interests when they filled in the survey in ways that possibly rendered insignificant the effect of social influence. One could also argue that social influence not only shapes and helps direct the activities and approaches of medical professionals, but also is itself influenced by socioeconomic circumstances or changes along with them. During the pandemic, social influence may therefore have focused on aspects or technologies more readily directed at managing the pandemic.

This study indicated that a higher perception of the availability of FC had a negative impact on the acceptance of AI-powered care pathways, which is contrary to previous studies [32]. This implies that if medical professionals perceive the FC in their medical organization—including training and technological resources—as better, they would be less likely to accept AI-powered care pathways. This could be explained by medical professionals perceiving good FCs as a workload increase due to additional trainings and technical tasks. Interestingly, it was found that gender moderates this relationship where identifying as female had a negative moderating effect, and identifying as male a positive moderating effect. This finding is in line with that reported by Haluza and Wernhart [33], who stated that there are gender divergences that are important to incorporate when formulating a new strategy for eHealth and telemedicine implementation.

However, narrowing in on the characteristics of the specific respondent groups reveals a nonrandom sample in terms of different medical professions (Multimedia Appendix 6); all nurses included in this study identified as female, whereas physicians largely identified as male. This nonrandom sample could have influenced the moderating effect of gender given the way in which tasks and responsibilities are distributed among these professional categories. Since nurses continue to perform more administrative work compared to physicians, they may view good FC as a constraint in that it may add to their already broad repertoire of tasks, whereas physicians may view FC as a helping tool particularly given their focus on MEPE [34]. In addition, the speed of technological advances in the work field requires continuous development of new skills, which might be more challenging to cope with for nurses owing to the highly varied nature of their tasks. Therefore, they may perceive the better FC more as a demand to keep up with the fast digitalization [35,36]. Besides the difference in job-specific tasks, the differences in professional identity and distribution of professional rewards based on the acquisition of new skills could also have an influence. Traditionally, a greater focus has been placed on the need for physicians to keep up to date with the latest clinical insights and approaches, and differences in professional status and social standing have often been derived on the basis of their frequent participation to such activities. In contrast, even though important transformations have taken place over the last decades in this sense, the main task of nurses is still seen by many as the provision of care, often understood as a quality that nurses somehow naturally possess rather than a set of skills that could be trained and fostered [37]. Thus, to the extent that such trainings may not lead to obvious professional rewards, nurses may see them more as a constraint and an imposition rather than as an opportunity. Furthermore, this result may have also been influenced by the broader and often gendered realities of nurses' lives, where family duties and other caring obligations outside their professional roles may prevent them from wanting or being able to take on new work roles and responsibilities. However, regional differences in the professional identity of nurses and physicians were found [38]. The limited sample in this study did not allow us to unambiguously prove if FC is indeed influenced by the profession or if it is mainly caused by gender. However, it is strongly suggested that both profession and gender play a role in how the perceived FC influences acceptance, so future studies should explore the relationship between these two variables and the underlying reasoning.

Strengths and Limitations

To our knowledge, this is the first study assessing the predictors of acceptance of AI-based technologies among medical professionals, thereby contributing to a poorly understood but increasingly relevant research area. A strength of this research was that it succeeded to identify significant relationships that influence acceptance. Another strength was the successful utilization of the UTAUT model and extensions of the model. This creates a foundation for future research in the acceptance of AI-based technologies. Furthermore, the quantitative nature of this study allows for more generalizable results and facilitates comparisons with future studies.

Some limitations were present in this study. Selection bias was unavoidable since the respondents participated voluntarily on the internet, which might have resulted in more individuals with an enthusiasm and interest about AI in health care. The selection bias could have been increased by the recruitment via the physicians' network. This might have resulted in more positive results since this network contains a lot of medical professionals with an interest in health technology.

The results revealed approximately 40% responses with missing values. Most health care professionals stopped the survey at the information page about AI-powered care pathways. This page required some reading and thus some effort to learn about AI-powered care pathways. Even though efforts made for the information provided about the AI-powered care pathways were succinct, the health care professionals may not have had the time to read these materials owing to the increased work pressure they experienced during the COVID-19 pandemic. Future iterations of this study should also interrogate respondents about the modalities through which they would be most successfully informed about these technologies when they are implemented. Visual or video materials might be more helpful when engaging with very busy professionals.

Furthermore, AI-powered care pathways are in the beginning of the implementation phase and therefore did not include the actual use behavior of AI-powered care pathways. This study could therefore not show if the acceptance is valid for predicting the actual use behavior.

Last, the used measures should be tested regarding their psychometric properties. Even though the constructs used in this study were mainly based upon validated models, the usefulness in the context of AI-powered care pathways needs

to be further investigated. In addition, new constructs were added and some constructs were adjusted, which requires further investigation. The internal consistency of the constructs was tested with the Cronbach α. Two constructs (FC and SIME) showed low internal consistency. FC was still included owing to the exploratory nature; we did consider lower Cronbach α values (<.5) since it adds critical information to the research. The internal consistency of SIME did not allow for its inclusion in deregression. This is a limitation since this study misses a potential predictor and it could still have contributed to nonzero amounts to the explained variance in the case of correlated regressors, which can be done by influencing other significant regressors.

Future Implications

This research should function as a foundation for future longitudinal research. Future research could identify if acceptance differs over adoption steps and when more awareness about the technology is present. This study was conducted in quite a premature stage where actual use is still limited.

Furthermore, future research should identify if the used model is applicable in different health care systems or in other regions of the world. Since this research was conducted in the Netherlands and included all type of medical organizations, variations between organizational cultures, differences in professional identity, and the difference in public opinion about AI were not taken into account. Insight into these differences could help develop adequate implementation strategies per region and organization.

Adaptations were made to make the model fit better to the research aim. Future studies should focus on further validating the model in the context of AI-based technologies, especially the construct with poor internal consistency.

Since performance expectancy was found as the strongest predictor for the acceptance of AI-powered care pathways, this should be high priority during implementation of AI-based health technologies. The added value of these technologies should be clearly communicated to the end users. PT was the second most influencing variable for the acceptance of AI-powered care pathways. Strategies on how to increase trust in AI-based technologies should therefore be formulated for successful adoption in health care. Even though trust is found to be an important facilitator for acceptance, future research should not only focus on how to increase trust but also what effect this trust has on the actual use, since studies found that people tend to overtrust and misinterpret the outcomes of AI-based decision support [39-41].

The quantitative nature of this study did not allow us to understand the medical professionals' reasoning underlying the found outcomes. Future qualitative studies are therefore recommended to understand how specific personality traits, the amount of understanding of AI-powered care pathways, or other contextual factors influence the acceptance of AI-based technologies.

Conclusions

This study sheds light on what factors have the largest impact on the acceptance of AI-powered care pathways among hospital staff and medical professionals. The model explained 75.4% of the variance in the behavioral intention. MEPE, NMPE, EE, PT, and PI were found to significantly influence behavioral intention where medical performance expectancy was found to have the largest impact. The moderator gender was found to significantly influence the relationship between facilitating conditions and acceptance. Since this study was conducted among Dutch medical professionals over a limited period of time and at a stage where the implementation of these technologies is still limited, follow-up surveys and multinational studies could further explore the predictors of acceptance of AI-powered care pathways over time and in different context.

Acknowledgments

We would like to thank Mirjam van der Steen and Katinka de Korte for their support during the study. We would also like to thank all the respondents who participated in our research.

Conflicts of Interest

VvB, LW, and DH are employed by DEARhealth, The Netherlands.

Multimedia Appendix 1
Survey items with the corresponding item sources.
[DOC File , 87 KB - formative_v6i6e33368_app1.doc]

Multimedia Appendix 2
Graphs and tables for assumption testing multiple linear regression.
[DOCX File , 171 KB - formative_v6i6e33368_app2.docx]

Multimedia Appendix 3
Pearson correlations between the variables.
[DOC File , 55 KB - formative_v6i6e33368_app3.doc]

Multimedia Appendix 4
Results from the multiple linear regression indicating the relationship between the predictor variables and the behavioral intention to use AI-powered care pathways.
[DOC File , 56 KB - formative_v6i6e33368_app4.doc]

Multimedia Appendix 5
Hayes' PROCESS regression matrix for the moderating effects on the relationships between the predictor variables and the behavioral intention to use AI-powered care pathways. The Coefficient, standard error and P-value of the interaction terms are shown.
[DOC File , 82 KB - formative_v6i6e33368_app5.doc]

Multimedia Appendix 6
Cross Table for gender x profession.
[DOC File , 48 KB - formative_v6i6e33368_app6.doc]

References

1. Johansen F, Loorbach D, Stoopendaal A. Exploring a transition in Dutch healthcare. J Health Organ Manag 2018 Oct 08;32(7):875-890 [FREE Full text] [doi: 10.1108/JHOM-07-2018-0185] [Medline: 30465486]
2. Mintz Y, Brodie R. Introduction to artificial intelligence in medicine. Minim Invasive Ther Allied Technol 2019 Apr;28(2):73-81. [doi: 10.1080/13645706.2019.1575882] [Medline: 30810430]
3. Tekkeşin. Artificial Intelligence in Healthcare: Past, Present and Future. Anatol J Cardiol 2019 Oct;22(Suppl 2):8-9 [FREE Full text] [doi: 10.14744/AnatolJCardiol.2019.28661] [Medline: 31670713]
4. Panesar A. Machine Learning and AI Ethics. In: Machine Learning and AI for Healthcare: Big Data for Improved Health Outcomes. Berkeley, CA: Apress; 2020:207-247.
5. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J 2019 Jun;6(2):94-98 [FREE Full text] [doi: 10.7861/futurehosp.6-2-94] [Medline: 31363513]
6. Reddy S, Fox J, Purohit M. Artificial intelligence-enabled healthcare delivery. J R Soc Med 2019 Jan;112(1):22-28 [FREE Full text] [doi: 10.1177/0141076818815510] [Medline: 30507284]
7. Higgins D, Madai V. From Bit to Bedside: A Practical Framework for Artificial Intelligence Product Development in Healthcare. Adv Intell Syst 2020 Jul 02;2(10):2000052 [FREE Full text] [doi: 10.1002/aisy.202000052]
8. Ahmad O, Stoyanov D, Lovat L. Barriers and pitfalls for artificial intelligence in gastroenterology: Ethical and regulatory issues. TIGE 2020 Apr;22(2):80-84 [FREE Full text] [doi: 10.1016/j.tgie.2019.150636]
9. Singh R, Hom G, Abramoff M, Campbell J, Chiang M, AAO Task Force on Artificial Intelligence. Current Challenges and Barriers to Real-World Artificial Intelligence Adoption for the Healthcare System, Provider, and the Patient. Transl Vis Sci Technol 2020 Aug;9(2):45 [FREE Full text] [doi: 10.1167/tvst.9.2.45] [Medline: 32879755]
10. Wang F, Kaushal R, Khullar D. Should Health Care Demand Interpretable Artificial Intelligence or Accept "Black Box" Medicine? Ann Intern Med 2020 Jan 07;172(1):59-60. [doi: 10.7326/M19-2548] [Medline: 31842204]
11. Selder A. Physician reimbursement and technology adoption. J Health Econ 2005 Sep;24(5):907-930. [doi: 10.1016/j.jhealeco.2005.03.004] [Medline: 15927287]
12. Laï MC, Brian M, Mamzer M. Perceptions of artificial intelligence in healthcare: findings from a qualitative survey study among actors in France. J Transl Med 2020 Jan 09;18(1):14 [FREE Full text] [doi: 10.1186/s12967-019-02204-y] [Medline: 31918710]
13. White K. Web-Based Questionnaire. In: Encyclopedia of Quality of Life and Well-Being Research. Dordrecht: Springer; 2014:301-311.
14. Venkatesh V, Morris M, Davis G, Davis F. User Acceptance of Information Technology: Toward a Unified View. MIS Quarterly 2003;27(3):425-478 [FREE Full text] [doi: 10.2307/30036540]
15. Shaw J, Rudzicz F, Jamieson T, Goldfarb A. Artificial Intelligence and the Implementation Challenge. J Med Internet Res 2019 Jul 10;21(7):e13659 [FREE Full text] [doi: 10.2196/13659] [Medline: 31293245]
16. Eckhardt A, Laumer S, Weitzel T. Who Influences Whom? Analyzing Workplace Referents' Social Influence on it Adoption and Non-Adoption. J Info Technol 2009 Mar 01;24(1):11-24. [doi: 10.1057/jit.2008.31]
17. Samhan B. Revisiting Technology Resistance: Current Insights and Future Directions. AJIS 2018 Jan 22;22. [doi: 10.3127/ajis.v22i0.1655]
18. Pavlou P. Consumer Acceptance of Electronic Commerce: Integrating Trust and Risk with the Technology Acceptance Model. Int J Electron Commer 2014 Dec 23;7(3):101-134 [FREE Full text] [doi: 10.1080/10864415.2003.11044275]
19. Simonson M, Maurer M, Montag-Torardi M, Whitaker M. Development of a Standardized Test of Computer Literacy and a Computer Anxiety Index. Journal of Educational Computing Research 1995 Jan 01;3(2):231-247 [FREE Full text] [doi: 10.2190/7chy-5cm0-4d00-6jcg]

20. Jussupow E, Spohrer K, Heinzl A, Link C. I am; We are - Conceptualizing Professional Identity Threats from Information Technology. 2018 Presented at: International Conference on Information Systems; December 13-16, 2018; San Francisco, CA.
21. Agarwal R, Prasad J. A Conceptual and Operational Definition of Personal Innovativeness in the Domain of Information Technology. Inf Syst Res 1998 Jun;9(2):204-215 [FREE Full text] [doi: 10.1287/isre.9.2.204]
22. Chen IJ, Yang KF, Tang FI, Huang CH, Yu S. Applying the technology acceptance model to explore public health nurses' intentions towards web-based learning: a cross-sectional questionnaire survey. Int J Nurs Stud 2008 Jun;45(6):869-878. [doi: 10.1016/j.ijnurstu.2006.11.011] [Medline: 17482191]
23. Adams K, Hean S, Sturgis P, Clark J. Investigating the factors influencing professional identity of first-year health and social care students. Learn Health Soc Care 2006 Jun;5(2):55-68 [FREE Full text] [doi: 10.1111/j.1473-6861.2006.00119.x]
24. Gliem J, Gliem R. Calculating, Interpreting, and Reporting Cronbach's Alpha Reliability Coefficient for Likert-Type Scales. 2003 Presented at: 2003 Midwest Research to Practice Conference in Adult, Continuing, and Community Education; 2003; Columbus, OH URL: http://hdl.handle.net/1805/344
25. Hayes AF. Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach. New York, NY: Guilford Publications; 2013.
26. Jun S, Plint A, Campbell S, Curtis S, Sabir K, Newton A. Point-of-care Cognitive Support Technology in Emergency Departments: A Scoping Review of Technology Acceptance by Clinicians. Acad Emerg Med 2018 May;25(5):494-507 [FREE Full text] [doi: 10.1111/acem.13325] [Medline: 28960689]
27. Tavares J, Oliveira T. Electronic Health Record Patient Portal Adoption by Health Care Consumers: An Acceptance Model and Survey. J Med Internet Res 2016 Mar 02;18(3):e49 [FREE Full text] [doi: 10.2196/jmir.5069] [Medline: 26935646]
28. van der Vaart R, Atema V, Evers A. Guided online self-management interventions in primary care: a survey on use, facilitators, and barriers. BMC Fam Pract 2016 Mar 09;17:27 [FREE Full text] [doi: 10.1186/s12875-016-0424-0] [Medline: 26961547]
29. Wu Y, Tao Y, Yang P. The use of unified theory of acceptance and use of technology to confer the behavioral model of 3G mobile telecommunication users. J Stat Manag Sys 2008 Sep;11(5):919-949 [FREE Full text] [doi: 10.1080/09720510.2008.10701351]
30. Alalwan A, Dwivedi Y, Rana N. Factors influencing adoption of mobile banking by Jordanian bank customers: Extending UTAUT2 with trust. Int J Info Manag 2017 Jun;37(3):99-110 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2017.01.002]
31. Hao H, Padman R. An empirical study of opinion leader effects on mobile technology implementation by physicians in an American community health system. Health Informatics J 2018 Sep;24(3):323-333 [FREE Full text] [doi: 10.1177/1460458216675499] [Medline: 30068266]
32. Bhattacherjee A, Hikmet N. Physicians' resistance toward healthcare information technology: a theoretical model and empirical test. Eur J Inf Syst 2017 Dec 19;16(6):725-737 [FREE Full text] [doi: 10.1057/palgrave.ejis.3000717]
33. Haluza D, Wernhart A. Does gender matter? Exploring perceptions regarding health technologies among employees and students at a medical university. Int J Med Inform 2019 Oct;130:103948. [doi: 10.1016/j.ijmedinf.2019.08.008] [Medline: 31442846]
34. Chang I, Hwang H, Hung W, Li Y. Physicians' acceptance of pharmacokinetics-based clinical decision support systems. Expert Syst Appl 2007 Aug;33(2):296-303 [FREE Full text] [doi: 10.1016/j.eswa.2006.05.001]
35. De Leeuw JA, Woltjer H, Kool RB. Identification of Factors Influencing the Adoption of Health Information Technology by Nurses Who Are Digitally Lagging: In-Depth Interview Study. J Med Internet Res 2020 Aug 14;22(8):e15630 [FREE Full text] [doi: 10.2196/15630] [Medline: 32663142]
36. McGrath M. The challenges of caring in a technological environment: critical care nurses' experiences. J Clin Nurs 2008 Apr;17(8):1096-1104. [doi: 10.1111/j.1365-2702.2007.02050.x] [Medline: 18321273]
37. ten Hoeve Y, Jansen G, Roodbol P. The nursing profession: public image, self-concept and professional identity. A discussion paper. J Adv Nurs 2014 Feb;70(2):295-309. [doi: 10.1111/jan.12177] [Medline: 23711235]
38. Kalisch B, Begeny S, Neumann S. The image of the nurse on the internet. Nurs Outlook 2007;55(4):182-188. [doi: 10.1016/j.outlook.2006.09.002] [Medline: 17678683]
39. Eiband M, Buschek D, Kremer A, Hussmann H. The Impact of Placebic Explanations on Trust in Intelligent Systems. 2019 Presented at: CHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow URL: https://doi.org/10.1145/3290607.3312787 [doi: 10.1145/3290607.3312787]
40. Ehsan U, Passi S, Liao Q, Chan L, Lee I, Muller M, et al. The Who in Explainable AI: How AI Background Shapes Perceptions of AI Explanations. arXiv. Preprint posted online July 28, 2021 [FREE Full text] [doi: 10.4324/9781003175063-5]
41. Wu E, Wu K, Daneshjou R, Ouyang D, Ho DE, Zou J. How medical AI devices are evaluated: limitations and recommendations from an analysis of FDA approvals. Nat Med 2021 Apr;27(4):582-584. [doi: 10.1038/s41591-021-01312-x] [Medline: 33820998]

Abbreviations

AI: artificial intelligence

AN: anxiety
BI: behavioral intention
EE: effort expectancy
FC: facilitating condition
IN: innovativeness
MEPE: medical performance expectancy
NMPE: nonmedical performance expectancy
PI: professional identity
PT: perceived trust
SI: social influence
SIME: social influence medical experts
SIPA: social influence patients
UTAUT: Unified Theory of Acceptance and Use of Technology

Edited by A Mavragani; submitted 04.09.21; peer-reviewed by J Tavares, JA Benítez-Andrades, D Valero-Bover; comments to author 21.12.21; revised version received 14.04.22; accepted 02.05.22; published 21.06.22.

Please cite as:
Cornelissen L, Egher C, van Beek V, Williamson L, Hommes D
The Drivers of Acceptance of Artificial Intelligence–Powered Care Pathways Among Medical Professionals: Web-Based Survey Study
JMIR Form Res 2022;6(6):e33368
URL: https://formative.jmir.org/2022/6/e33368
doi:10.2196/33368
PMID:35727614

Original Paper

UK Adults' Exercise Locations, Use of Digital Programs, and Associations with Physical Activity During the COVID-19 Pandemic: Longitudinal Analysis of Data From the Health Behaviours During the COVID-19 Pandemic Study

Verena Schneider[1], MSc; Dimitra Kale[1,2], PhD; Aleksandra Herbec[1,3], PhD; Emma Beard[1,2], PhD; Abigail Fisher[1], PhD; Lion Shahab[1,2], PhD

[1]Department of Behavioural Science and Health, University College London, London, United Kingdom

[2]SPECTRUM Research Consortium, Edinburgh, United Kingdom

[3]Clinical, Educational and Health Psychology, University College London, London, United Kingdom

Corresponding Author:
Verena Schneider, MSc
Department of Behavioural Science and Health
University College London
1–19 Torrington Place
London, WC1E 6BT
United Kingdom
Phone: 44 20 7679 1720
Email: verena.schneider.19@ucl.ac.uk

Abstract

Background: Digital physical activity (PA) program use has been associated with higher PA guideline adherence during COVID-19 pandemic confinements. However, little is known longitudinally about exercise locations (inside vs outside the home environment), digital program use, and their associations with moderate-to-vigorous PA (MVPA) and muscle-strengthening activities (MSAs) during the pandemic.

Objective: The aims of this study were to assess the relationship between exercise location and use of digital programs with PA guideline adherence during the COVID-19 pandemic, describe how individuals exercised inside and outside of their home environments, and explore which sociodemographic and contextual factors were associated with exercise locations and digital PA program use.

Methods: Active UK adults (N=1938) who participated in the 1-month follow-up survey of the Health Behaviours During the COVID-19 Pandemic (HEBECO) study (FU1, June-July 2020) and at least one more follow-up survey (FU2, August-September; FU3, November-December 2020) reported exercise locations and types of exercises inside and outside their homes, including digital programs (online/app-based fitness classes/programs), MVPA, and MSA. Generalized linear mixed models were used to assess associations of exercise location and digital PA program use with PA guideline adherence (MVPA, MSA, full [combined] adherence), and predictors of exercise location and digital program use.

Results: As the pandemic progressed, active UK adults were less likely to exercise inside or to use digital PA programs compared with periods of initial confinement: 61% (95% CI 58%-63%; weighted n=1024), 50% (95% CI 48%-53%; weighted n=786), and 49% (95% CI 46%-51%; weighted n=723) performed any exercise inside their homes at FU1, FU2, and FU3, respectively. At FU1, FU2, and FU3, 22% (95% CI 21%-25%; weighted n=385), 17% (95% CI 15%-19%; weighted n=265), and 16% (95% CI 14%-18%; weighted n=241) used digital PA programs, respectively. Most participants who exercised inside already owned indoor equipment, used digital PA programs, or had their own workout routines, whereas MVPA and gentle walking were the most common exercise types performed outside the home. Being female, nonwhite, having a condition limiting PA, indoor exercising space, a lower BMI, and living in total isolation were associated with increased odds of exercising inside the home or garden compared with outside exercise only. Digital PA program users were more likely to be younger, female, highly educated, have indoor space to exercise, and a lower BMI. While exercising inside was positively associated with MSA and exercising outside was positively associated with MVPA guideline adherence, both inside (vs outside only) and outside (vs inside only) activities contributed to full PA guideline adherence (odds ratio [OR] 5.05, 95% CI 3.17-8.03 and OR 1.89, 95% CI 1.10-3.23, respectively).

Digital PA program use was associated with a higher odds of MSA (OR 3.97-8.71) and full PA (OR 2.24-3.95), but not with MVPA guideline adherence.

Conclusions: During the COVID-19 pandemic, full PA guideline adherence was associated with exercising inside and outside of one's home environment and using digital PA programs. More research is needed to understand the reach, long-term adherence, and differences between digital PA solutions.

(JMIR Form Res 2022;6(6):e35021) doi:10.2196/35021

KEYWORDS

pandemic; physical activity; longitudinal; United Kingdom; digital health; tele-exercise; moderate-to-vigorous physical activity; muscle-strengthening activity; COVID-19; home-based exercise; exercise; telemedicine; longitudinal; health behavior; behavior; data

Introduction

Insufficient physical activity (PA) and sedentary behavior are among the leading risk factors for premature mortality and chronic conditions, and present a global public health concern [1,2]. The World Health Organization (WHO) recommends 150 weekly minutes of moderate-to-vigorous PA (MVPA) and two weekly sessions of muscle-strengthening activities (MSAs) for adults [2]. However, approximately one-third of the English population aged 16 years or over did not meet MVPA guidelines in 2019 [3], and the prevalence of MSA guideline adherence is reported to be as low as 10%-30% across countries [4]. The economic costs of insufficient PA for the National Health Service England are estimated at £450 million (~US $568 million) a year [5]. The COVID-19 pandemic confinements have been linked to worldwide declines in PA levels [6] and changes in individuals' exercise habits [7]. To mitigate the impact of the COVID-19 pandemic and to inform responses to future pandemics, it is important to understand what helps individuals to exercise sufficiently during restrictions and at different phases of the pandemic, and whether digital PA programs can support this activity.

Population-level negative impacts of the initial pandemic confinements on PA and sedentary behavior have been reported in multiple observational studies [6,8-10]. In England, the proportion of active individuals aged 16 years or older dropped by 7.1% to 58.2% during the first UK lockdown [11]. However, despite the frequently reported population-level decline, the impact was not equal across demographic groups, as some individuals were able to maintain or even increase PA levels [12]. The groups most strongly impacted by PA declines were individuals with higher baseline PA levels [10,13,14], and those who were employed [12], in lower socioeconomic positions [3,11,12,15-17], female [16-18], nonwhite [3,11,19], living alone [17], with higher BMI [17], or living with a health condition [3,17,20]. Although some studies reported negative impacts for older age groups [16,18], others found that younger adults were most strongly affected [3,10,12,21]. In addition, having access to space to exercise, at home or within the neighborhood [17,19], and living with others rather than alone [17] have been identified as protective factors against declines or low PA levels.

As exercise contexts changed due to the pandemic, PA habits have likely been disrupted [7]. Although this may pose a risk to healthy habits, new opportunities and contexts may also promote the uptake of beneficial behaviors. However, UK evidence on long-term PA changes throughout the pandemic is scarce and conflicting. Repeated cross-sectional data taking into account prepandemic seasonal activity trends suggest a partial recovery of the pandemic impact during the summer and autumn of 2020 compared to the first lockdown [11]. Longitudinal data suggest positive long-term changes in PA behaviors from lockdown into easing of restrictions [22] or continued PA declines into the autumn/winter 2020 [23].

While seasonal [24,25] and pandemic-specific barriers may affect outdoor PA and opportunities to exercise in gyms, leisure facilities, and organized sports, engagement in PA at home may be associated with fewer declines in PA. For example, exercising in one's home or driveway was associated with higher MVPA in a cross-sectional study with US adults in the first 2 months of COVID-19 restrictions [19]. Different exercise locations lend themselves to different exercise types due to the availability of equipment and space. Thus, exercise locations may be differently related to MVPA and MSA. For example, exercises inside one's home may have higher components of MSA due to strengthening exercises not requiring a lot of space. By contrast, exercises outside one's home may have higher components of MVPA, as individuals are more likely to engage in aerobic activities such as brisk walking, running, or cycling. In the initial UK lockdown, half of those who exercised reported substantial changes in the form of exercise (none or some of the same exercises) [26]; however, little is known on how or if the forms of exercises changed over different phases of the pandemic and across seasons.

Additionally, the increasingly prevalent use of digital technologies such as web-based and smartphone-based programs and services, including apps, may provide additional support for the engagement in, and maintenance of, PA behavior. Studies conducted during the initial confinements suggest that users of digital support such as PA apps or online platforms were more likely to meet recommended PA guidelines [27,28] and less likely to experience a decrease in PA [16,29]. However, none of these studies used data from different phases of the pandemic beyond the initial confinements. Further, only one cross-sectional study conducted during the initial lockdown of the COVID-19 pandemic in Australia included measures of MSA [28]. This study reported that 39.5% of adults used some form of digital PA platform, of which streaming services (eg, YouTube, Instagram, and Facebook) and facilitated live or recorded online classes (eg, via Zoom) were the most frequently

reported (42% and 31%, respectively). Compared with nonusers, digital PA platform users were 2.7-times more likely to meet WHO recommendations for both MVPA and MSA [28].

Although observational studies are limited regarding causal conclusions, describing naturally occurring behavior and associated factors in observational real-world studies can help hypothesis generation for further research and intervention development. By identifying factors associated with exercise locations and digital program use, different target groups, and potential barriers, facilitators, and risks relating to feasibility, reach, and adherence over time can be identified. Research conducted before and during the pandemic identified users of health apps or digital platforms as more likely to be female [28,30], younger [31,32], frequent smartphone users [30], with higher education and income [31], with a chronic condition [32], employed, and without home or caring duties [28]. However, little is known to date about factors associated with exercise locations and digital program use or about their associations with MVPA and MSA levels, and the changes in the ways of exercising across different phases of the pandemic.

To address this gap, the primary aim of this study was to assess the relationship between exercise location and use of digital PA programs with MVPA, MSA, and full guideline adherence among active UK adults during the COVID-19 pandemic (between May and December 2020). Secondary aims were to describe changes over the course of the COVID-19 pandemic in exercise location, how individuals exercised inside and outside of their homes, and to explore which sociodemographic and contextual predictors were associated with the choice of exercise locations and digital PA program use.

Thus, this study sought to answer the following research questions (RQs) using longitudinal data collected during the COVID-19 pandemic: (RQ1) What were the differences in exercise location and the ways of exercising in June-July, August-September, and November-December 2020 during the COVID-19 pandemic? (RQ2) What demographic and contextual factors were associated with exercise location and use of digital PA programs in June-July, August-September, and November-December 2020 in active UK adults during the COVID-19 pandemic? (RQ3) What was the association of exercise location and digital PA program use with PA guideline adherence (MVPA, MSA, and combined) in active UK adults in June-July, August-September, and November-December 2020 during the COVID-19 pandemic?

Methods

Design

This study analyzed longitudinal data from the Health Behaviours During the COVID-19 Pandemic (HEBECO) study [33]. Baseline data collection took place between April 23 and June 14, 2020, during the first UK-wide lockdown, with follow-up (FU) questionnaires sent out at 1 month (FU1; June-July 2020, lockdown/some lifts of restrictions), 3 months (FU2; August-September 2020, fewer restrictions), and 6 months (FU3; November-December 2020, country-specific lockdown/restrictions). Full details of the pandemic context for each data collection phase are described in Multimedia Appendix 1. As data on the key outcome variables were only collected as part of the core follow-up surveys, this study only included the three time points at FU1, FU2, and FU3. This study was preregistered on Open Science Framework [34] (see Multimedia Appendix 2 for changes to the protocol).

Ethics Approval

Ethical approval was granted by University College London (UCL) Research Ethics Committee at UCL Division of Psychology and Language Sciences (CEHP/2020/579).

Recruitment

Participants were recruited into the HEBECO survey through various channels such as paid advertisements, social media, charities, and partner organizations [33]. For the purposes of this study, non-UK residents or those who were completely physically inactive at all time points (MVPA=0 and MSA=0) were excluded. The latter criterion was a methodological consideration to reduce a risk of bias from including inactive participants in models of associations. The current survey was not set up to identify any previous download of digital programs but rather specifically asked about whether participants engaged in PA had used digital programs to do so. Since it is logically impossible to be simultaneously inactive while exercising using digital PA programs, including inactive participants would create a meaningless or inflated association between program use and PA guideline adherence. Thus, to ensure the results' internal validity, eligible participants needed to have self-reported any MSA or MVPA for at least one time point.

Measures

Outcome Measures

Full details and wording of measures can be found in the protocol [34] and Multimedia Appendix 3. All outcome variables were collected repeatedly during FU1, FU2, and FU3.

WHO PA guideline adherence (MVPA, MSA, and combined) was reported using validated questions based on the Behavioral Risk Factor Surveillance System 2015 [35,36]. MVPA was assessed by asking participants (1) how many times on average per week they had done a minimum of 15 minutes of MVPA (eg, brisk walking, jogging, dancing, cycling) and (2) how long (in minutes) an average session had been in the past month. Weekly average MVPA was defined as the product of these two variables. MSA was assessed by asking participants how many days per week on average they had done strength training in the past month. Three binary outcome variables were created to indicate individuals who met WHO MVPA recommendations (MVPA≥150 minutes/week) versus not, individuals who met WHO MSA recommendations (MSA≥2 sessions/week) versus not, and individuals who met both full recommendations versus not. The reported 2-week retest reliabilities of the MVPA and MSA measures are considered substantial (Cohen κ=0.67 and κ=0.85, respectively) and concurrent validities with activity logs are considered moderate (κ=0.41 and κ=0.52, respectively) [35].

Exercise location was assessed by asking participants who indicated engaging in any level of MVPA or MSA whether they

had been exercising inside, outside, or both inside and outside their house/garden (items generated by the HEBECO study team). To account for the effects of doing any exercise either inside or outside one's home environment with mutually exclusive categories, these were dichotomized into two variables: (1) any activity in the home environment versus only outside and (2) any activity outside of the home environment versus only inside.

Type of exercise was assessed by asking participants who indicated exercising inside or outside their home environments "What exercises are you usually doing inside/outside your house/garden?" Multiple answers were possible and were combined by creating the dichotomous variables (1) gentle walking (vs not), (2) MVPA activities such as any brisk walking/alternate walking-running/running/cycling/swimming (vs not), (3) any team/racket sports (vs not), (4) weightlifting (vs not), (5) online/app-based fitness classes/program (vs not), and (6) other (vs not) for activities outside the home environment. Activities inside the home environment were (1) exercise DVD (vs not), (2) online/app-based fitness classes/program (vs not), (3) using indoor exercise equipment that I already had (vs not), (4) using indoor exercise equipment that I bought/borrowed during COVID-19 (vs not), (5) doing bodyweight exercises without using an online class or app (ie, your own workout; vs not), and (6) other (vs not). Open-text responses on the "other" category were included in the FU1 and FU2 surveys only, precluding a systematic coding of these answers.

Use of digital PA programs was a dichotomous variable indicating individuals who had reported exercising using any "online/app-based fitness classes/program" (inside or outside) versus not.

Sociodemographic Predictors

Sociodemographic predictors collected at baseline were gender (female, others), age (<35 years as reference, 35-64 years, and ≥65 years), ethnicity (white, other), education (≥16 years, <16 years), health condition limiting PA (yes, no), and country of UK residence (England, other countries). Space to exercise comfortably inside one's home or garden was only assessed at FU2 and FU3 and therefore dichotomized into no space on at least one time point (reference category vs all others).

Time-Variant Predictors

Repeatedly measured predictors were employment (full/part-time vs others), COVID-19–induced isolation (total vs some, general, no isolation [reference]), BMI (continuous), perceived risk of COVID-19 to one's health (major-significant, lower), smoking (current, not), and alcohol consumption per week (>14, ≤14 units; [37]). Time was measured in months to account for the unequal time intervals between measurement points (1, 3, and 6 months) and centered at zero (0, 2, and 5). In addition to the key variables exercise location and digital PA program use, as described in the outcome measures, a time×exercise location interaction term was created to assess any differences in associations between location and PA guideline adherence over time.

Statistical Analysis

Descriptive analyses and assumption checks were performed in SPSS 27.0. Weighted data were used to account for nonrandom sampling when describing the sample, and their exercise locations and types (RQ1). Weights account for population proportions of gender, age, ethnicity, household income, and country [38]. Descriptive statistics were calculated to describe the sample on key demographic and study variables. The analytic sample and participants lost to follow-up since baseline were compared on baseline characteristics using *t*-tests for continuous variables and χ^2 tests for categorical variables. Descriptive statistics on exercise location, and activities inside and outside the home environment were computed as percentages per wave with cross-sectionally complete data.

RQ2 and RQ3 were assessed by running generalized linear mixed models (GLMMs) with dichotomous outcomes in R using the lme4 package [39]. First, linearity of the continuous predictors with the log of the outcome were checked by entering the predictor and its interaction effect with the log of itself into the model. According to Field [40], a significant interaction effect indicates a problem with linearity. As the continuous age variable was violating the linearity assumption, the categorical variable was used throughout, as specified in the Measurement section. Second, checks were run to ensure that multicollinearity was not present, which included inspection of the correlation matrix for correlations ≥0.8 and the calculation of variance inflation factors. Any variance inflation factor≥10 would have been considered problematic [40].

For RQ2, GLMMs were run to assess predictors of exercise location (exercising inside vs only outside and exercising outside vs only inside one's home environment) and digital PA program use (binary logistic mixed model, reference: none). Repeated-measures (level-1) variables were nested within participants (level-2) and grand mean–centered. Univariate and fully adjusted models with random intercepts were run by including all the above-listed time-variant and -invariant predictors (except smoking and alcohol consumption). Similarly, GLMMs were run for RQ3, predicting MVPA, MSA, and full PA guideline adherence. First, unadjusted and adjusted models with time, the key predictors (exercise location, use of digital classes), and the interaction term with time were run. Second, models were fully adjusted for the remaining predictors. Sensitivity analyses were performed on a data set including only participants with complete data in all waves (FU1, FU2, and FU3). Significance thresholds in unadjusted models were Benjamini-Hochberg–adjusted to account for family-wise error [41].

In the absence of significant effects, Bayes factors (BFs) were computed using an online calculator [42] to distinguish insensitive data ($1/3<|BF|\leq 3$) from an absence of an effect ($|BF|<1/3$). Absolute BFs>3 were considered as moderate relative evidence for an effect. Based on associations of digital platform use with PA guideline adherence reported by Parker et al [28], half-normal distributions (eg, one-sided tests) with hypothesized effects of odds ratio (OR)=2.0 (MVPA), OR=3.3 (MSA), and OR=2.7 (full guidelines) were specified for RQ3.

Results

Sample Characteristics

Of the 2992 UK adults who participated in the baseline survey, 2363 (78.98%) started the FU1 survey. Of these, 14 (0.59%) moved out of the United Kingdom and 253 (10.71%) did not participate in any further follow-up. An additional 158 were excluded due to complete physical inactivity. Thus, the final analytic sample consisted of 1938 UK adults who provided a total of 5429 observations. When applying baseline weights to account for nonrandom sampling, the analytic sample was n=1680 (sample lost to follow-up n= 680).

Table 1 presents the baseline characteristics of the analytic sample and participants lost from baseline to FU1 (for unweighted estimates see Multimedia Appendix 4). The weighted analytic sample consisted of a higher proportion of individuals older than 64 years, of white ethnicity, and higher education, and a lower proportion of individuals living in total isolation, smokers, and individuals adhering to full PA guidelines at baseline. Further, the analytic sample had a significantly higher BMI than that of participants lost to follow-up.

Table 1. Baseline characteristics of the analytic sample and participants lost to follow-up (weighted population estimates, baseline weights).

Characteristic	Analytic sample (n=1680)	Sample lost to follow-up (n=680)	*P* value
Age (years), weighted n (%)			<.001
<35	264 (15.7)	272 (40.0)	
35-64	1087 (64.7)	349 (51.3)	
>64	329 (19.6)	59 (8.7)	
Female, weighted n (%)	868 (51.7)	355 (52.2)	.81
White ethnicity, weighted n (%)	1529 (91.0)	558 (82.1)	<.001
16+ years of education, weighted n (%)	1146 (68.2)	432 (63.5)	.03
Employed, weighted n (%)	813 (48.4)	337 (49.6)	.62
Condition limiting PA[a], weighted n (%)	240 (14.4)	107 (16.2)	.27
Living in England, weighted n (%)	1419 (84.5)	552 (81.2)	.05
Total isolation, weighted n (%)	93 (5.6)	66 (10.0)	<.001
High perceived risk from COVID-19, weighted n (%)	432 (25.9)	151 (22.8)	.11
Smoker, weighted n (%)	325 (19.4)	267 (39.3)	<.001
High alcohol consumption, weighted n (%)	288 (18.4)	127 (22.0)	.06
Meeting WHO[b] PA recommendations at baseline, weighted n (%)			
MVPA[c]	676 (42.0)	248 (41.2)	.73
MSA[d]	499 (31.0)	182 (30.1)	.70
Both	265 (16.5)	121 (20.1)	.046
BMI, mean (SD)	26.5 (4.9)	25.6 (5.2)	<.001

[a]PA: physical activity.

[b]WHO: World Health Organization.

[c]MVPA: moderate-to-vigorous physical activity.

[d]MSA: muscle-strengthening activity.

RQ1: Exercise Locations, Use of Digital PA Programs, and PA Behaviors

Table 2 presents descriptive data on exercise locations, use of digital PA programs, and WHO guideline adherence at the three time points (for unweighted estimates see Multimedia Appendix 5). Across time, most participants exercised only outside or both inside and outside their home environments (68%-74%), whereas fewer individuals exercised inside their home environments only (15%-18%). The proportions of individuals who did any exercise inside their home environments was 61% (95% CI 58%-63%), 50% (95% CI 48%-53%), and 49% (95% CI 46%-51%) at FU1 (June-July 2020), FU2 (August-September 2020), and FU3 (November-December 2020), respectively. While over one-fifth (23%, 95% CI 21%-25%) of active adults used digital PA programs at FU1, the proportions were 17% (95% CI 15%-19%) and 16% (95% CI 14%-18%) at FU2 and FU3, respectively. In addition, 18% of participants (95% CI 16%-20%) adhered to the full WHO guidelines at FU1, 13% (95% CI 12-15%) at FU2, and 13% (95% CI 12%-15%) at FU3.

The most frequently reported ways of exercising inside one's home environment were using already owned indoor equipment, digital PA programs, one's own workout, or other (Figure 1).

Other exercise types as indicated in open-text responses in FU1 and FU2 included gardening, do-it-yourself (DIY) activities, physiotherapy exercises, playing with children, Pilates, yoga, stretching, gymnastics, dancing, martial arts, and private personal trainer sessions. Less frequent were use of DVD or bought/borrowed equipment. The highest relative frequency of reporting the use of own indoor equipment and digital PA programs was observed at FU1 (June-July 2020).

MVPA was the most frequently reported exercise type outside of the home environment, followed by gentle walking (Figure 2). Frequent open-text responses in the "other" category in FU1 and FU2 included horse riding or looking after horses, dog agility, water sports, climbing/hiking, golf, martial arts, the return to the gym, and various group and private trainer sessions. The highest relative frequency of MVPA was reported during FU2 (August-September 2020), a time of fewer restrictions, which also saw the highest relative frequency of team and racket sports compared with other time points.

Table 2. Exercise locations, use of digital programs, and meeting of World Health Organization (WHO) physical activity (PA) recommendations at follow-up 1 (FU1), follow-up 2 (FU2), and follow-up 3 (FU3).

Measure	FU1 (n=1725)[a], weighted n (%)	FU2 (n=1587)[a], weighted n (%)	FU3 (n=1535)[a], weighted n (%)
Exercise location[b]			
Inside home environment	312 (18.3)	238 (15.3)	238 (16.1)
Outside home environment	526 (30.9)	614 (39.4)	522 (35.3)
Both inside and outside	721 (41.9)	548 (35.1)	485 (32.8)
Use of digital PA programs[b]	385 (22.6)	265 (17.0)	241 (16.3)
Meeting WHO recommendations[c]			
MVPA[d]	750 (44.0)	624 (39.9)	582 (39.3)
MSA[e]	538 (31.6)	467 (29.9)	409 (27.7)
Both	301 (17.7)	206 (13.2)	199 (13.5)

[a]Note that the n values differ from those in Table 1 due to different weights being applied (FU1, FU2, and FU3 weights vs baseline weights). Percentages are valid percentages (ie, excluding missingness).

[b]Missingness: FU1=24, FU2=27, FU3=56. Total N in exercise location includes active participants who dropped to inactivity at a certain time point (neither exercised inside nor outside the home environment); hence, percentages do not add up to 100%.[c]Missingness: FU1=22, FU2=24, FU3=56.

[d]MVPA: moderate-to-vigorous physical activity.

[e]MSA: muscle-strengthening activity.

Figure 1. Exercise types inside one's home environment (weighted population estimates). FU: follow-up.

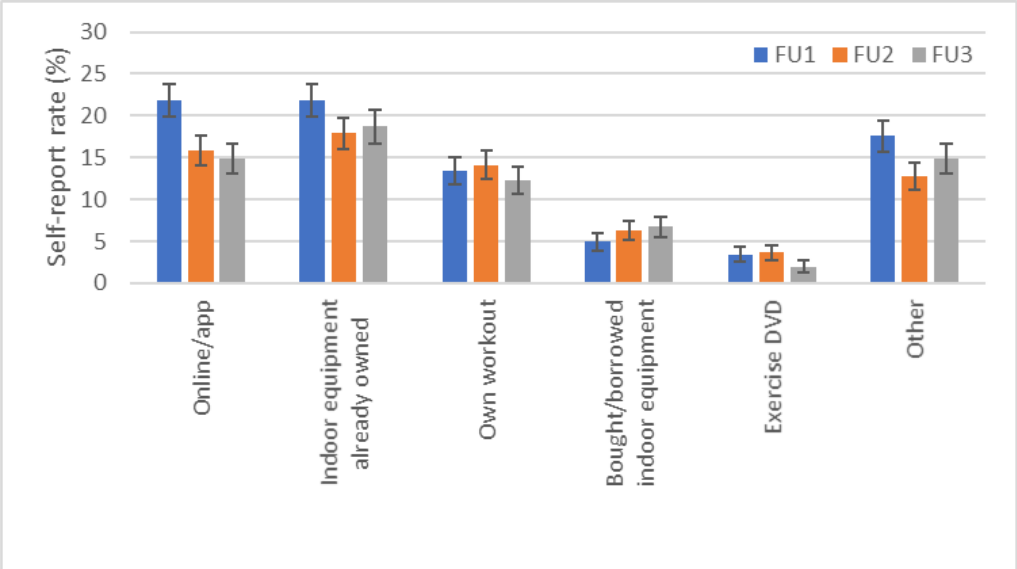
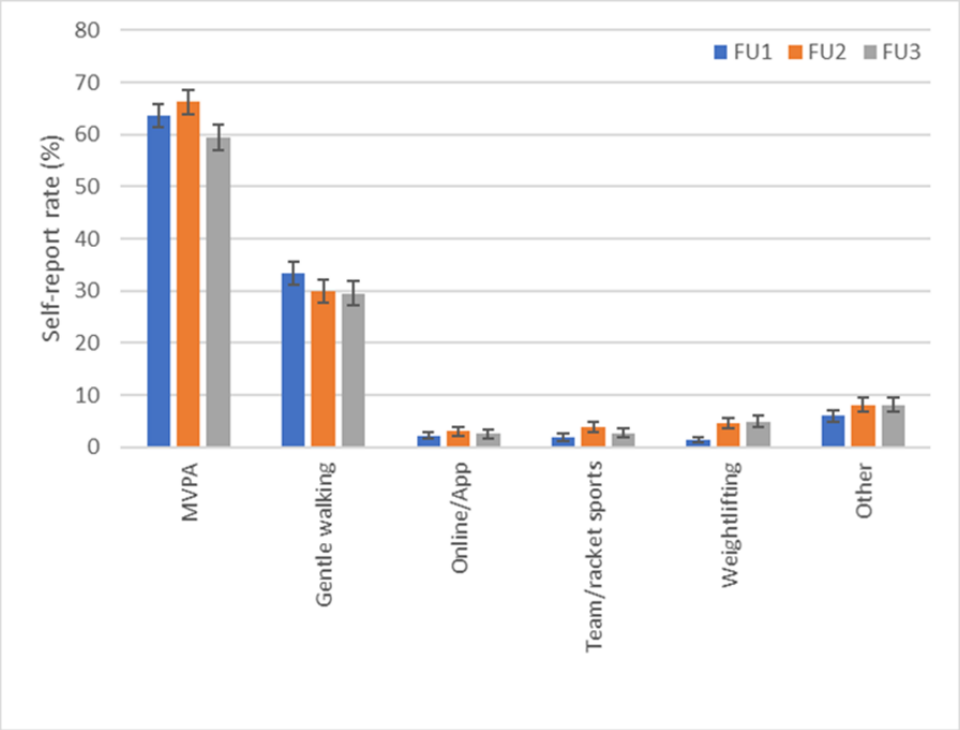


Figure 2. Exercise types outside one's home environment (weighted population estimates). FU: follow-up.



RQ2: Predictors of Exercise Locations and Digital PA Program Use

Factors associated with increased odds of exercising inside the home environment were female gender, having a condition limiting PA, having indoor space, and living in total isolation, whereas being in the 35-64–year age group (vs <35 years) and white ethnicity were associated with decreased odds of exercising inside (Table 3; for unadjusted analyses see Multimedia Appendix 6). A 1-point increase in BMI was associated with a 3% decrease in the odds of exercising inside versus outside the home environment only. However, the associations with age and BMI were not robust in the complete case analysis (Multimedia Appendix 7).

Associated factors with increased odds of exercising outside of the home environment were older age (>64 vs <35 years), higher education, having no condition limiting PA, not having indoor space to exercise, a lower perceived risk from COVID-19, and not living in total isolation. While the odds of exercising inside were significantly reduced at both FU2 and FU3 compared with FU1, time was not a significant predictor of the odds of exercising outside the home environment.

The odds of using digital PA programs decreased at FU2 and FU3 compared with FU1. Associated factors with increased odds of using digital PA programs were younger age, female gender, higher education, and indoor space. A 1-point increase in BMI was associated with an 8% decrease in the odds of using digital PA programs. Complete case analyses replicated these findings, except for education, which was nonsignificant (Multimedia Appendix 7).

Table 3. Fully adjusted generalized linear mixed model estimates of the predictors of exercising inside (vs only outside) or outside (vs only inside) the home environment, and of using digital physical activity (PA) programs (vs not) at follow-up 1 (FU1), follow-up 2 (FU2), and follow-up 3 (FU3).

Predictor	Exercising inside[a]		Exercising outside[a]		Digital PA program use[b]	
	OR[c] (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Time[d] (reference: FU1)						
FU2	0.51 (0.42-0.62)	<.001	1.23 (0.89-1.70)	.20	0.47 (0.36-0.59)	<.001
FU3	0.57 (0.47-0.70)	<.001	1.13 (0.81-1.58)	.46	0.50 (0.39-0.64)	<.001
Age (years) (reference: <35 years)						
35-64	0.70 (0.49-1.00)	.048	1.60 (0.78-3.28)	.20	0.24 (0.14-0.40)	<.001
>64	0.84 (0.54-1.31)	.45	3.44 (1.36-8.71)	.009	0.10 (0.05-0.21)	<.001
Female gender (reference: all other)	1.34 (1.03-1.75)	.03	1.17 (0.68-2.04)	.57	6.91 (4.46-10.71)	<.001
White ethnicity (reference: nonwhite)	0.46 (0.25-0.84)	.01	2.91 (0.89-9.48)	.08	0.51 (0.23-1.15)	.11
High education (reference: <16 years)	0.86 (0.58-1.27)	.45	2.89 (1.31-6.38)	.009	2.56 (1.37-4.76)	.003
Condition limiting PA (reference: none)	1.65 (1.12-2.45)	.01	0.31 (0.14-0.69)	.004	0.75 (0.41-1.34)	.33
England (reference: all other UK countries)	1.16 (0.82-1.65)	.41	0.67 (0.31-1.44)	.31	1.12 (0.65-1.92)	.68
Indoor space (reference: none)	6.12 (4.60-8.13)	<.001	0.56 (0.33-0.95)	.03	12.79 (8.30-19.69)	<.001
Employed (reference: not employed)	1.05 (0.82-1.34)	.72	1.52 (0.96-2.42)	.07	1.33 (0.94-1.87)	.11
BMI	0.97 (0.95-0.99)	.02	0.98 (0.93-1.03)	.37	0.92 (0.89-0.96)	<.001
High perceived risk of COVID-19 (reference: low)	1.30 (1.00-1.69)	.05	0.48 (0.30-0.77)	.002	0.95 (0.65-1.37)	.77
Total isolation (reference: not)	5.08 (2.18-11.82)	<.001	0.01 (0.00-0.03)	<.001	1.36 (0.63-2.95)	.43

[a]N=4492 observations, n=1772 individuals.

[b]N=4865 cases, n=1780 individuals.

[c]OR: odds ratio.

[d]Time violated the linearity assumption and was thus entered as a categorical variable.

RQ3: Associations with PA Guideline Adherence

The odds of full guideline adherence decreased over time in active adults in the unadjusted analyses. However, the odds were attenuated when adjusting for the key predictors of location and digital program use (Multimedia Appendix 8), and were further attenuated to nonsignificance when adjusting for all remaining predictors in the analysis using the predictor exercising inside versus only outside the home environment (Table 4). Similarly, a significant decrease in the odds of adhering to MSA guidelines over time was attenuated when adjusting for the key predictors.

When fully adjusted for all other predictors, active adults exercising inside versus only outside their home environment had 5-times the odds of adhering to full PA guidelines (Table 4; for full tables with covariate estimates see Multimedia Appendix 9). Although they had 0.5-times reduced odds of adhering to MVPA guidelines, they had 9.7-times increased odds to adhere to MSA guidelines compared with adults who only exercised outside the home environment. The significant interaction between exercising inside and time indicated that the associations of exercising inside with MSA and full PA guideline adherence significantly differed over time, although this was not robust in complete case analyses (see Multimedia Appendix 10). Associations of exercising inside (vs outside the home environment only) with MSA and full PA guideline adherence were stronger at FU1 (OR 6.7, 95% CI 4.8-9.4 and OR 3.7, 95% CI 2.5-5.6, respectively) compared with FU2 (OR 4.5, 95% CI 3.3-6.0 and OR 2.9, 95% CI 2.0-4.1, respectively) and FU3 (OR 4.2, 95% CI 3.0-5.8 and OR 2.4, 95% CI 1.6-3.6, respectively; Multimedia Appendix 11).

Users of digital PA programs had 4-times the odds of adhering to MSA and 2.2-times the odds of adhering to full PA guidelines than active adults who did not use these programs. The association with MVPA was not significant. The BF of 0.37 indicated inconclusive evidence, although the OR<1 in the complete case analysis indicated an absence of an effect (BF=0.14; Multimedia Appendix 10).

When replacing the key variable exercising inside with exercising outside (vs inside the home environment only), exercising outside was associated with 4.4-times the odds of MVPA and 1.9-times the odds of full guideline adherence in active adults compared to those who only exercised inside their homes (Table 5; for unadjusted analyses, full tables with covariate estimates, and complete case analyses see Multimedia Appendices 8, 12, and 13 respectively). Further, people who exercised outside had 0.4-times the odds of adhering to MSA guidelines compared with those who exercised inside their home environment only. These associations did not significantly differ over time as indicated by the nonsignificant interaction. Users of digital PA programs were 8.7-times more likely to adhere to MSA and were 4-times more likely to adhere to full PA

guidelines. Again, there was an absence of effect on MVPA guideline adherence (BF=0.14). Although time was associated with significantly decreased odds of meeting MVPA when entering exercising inside (vs outside only) as a predictor (Table 4), it was associated with decreased odds of meeting full guidelines in the model including exercising outside (vs inside the home environment only) as a predictor (Table 5).

Table 4. Generalized linear mixed model estimates predicting meeting moderate-to-vigorous activity (MVPA), muscle-strengthening activity (MSA), and full recommendations (vs not) at follow-up 1, 2, and 3; for key predictor exercising inside (vs outside the home environment only).

Predictor	MVPA[a]		MSA[a]		Full PA[b] recommendations[a]	
	OR[c] (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Time	0.81 (0.66-0.99)	.04	1.02 (0.80-1.30)	.86	0.84 (0.63-1.12)	.23
Exercising inside (reference: outside only)	0.54 (0.39-0.73)	<.001	9.70 (6.52-14.44)	<.001	5.05 (3.17-8.03)	<.001
Use of digital PA programs (reference: not)	1.10 (0.83-1.45)	.50	3.97 (2.92-5.38)	<.001	2.24 (1.60-3.13)	<.001
Time×location interaction	0.99 (0.91-1.08)	.85	0.87 (0.78-0.97)	.01	0.87 (0.76-0.99)	.04

[a]N=4439 observations, n=1769 individuals. Models fully controlled for age, gender, ethnicity, education, condition limiting PA, country, indoor space, employment, BMI, perceived risk of COVID-19, isolation status, smoking, and alcohol consumption. Bayes factor for nonsignificant associations with digital PA program use was 0.37 (MVPA).

[b]PA: physical activity.

[c]OR: odds ratio.

Table 5. Unadjusted and fully adjusted generalized linear mixed model estimates predicting meeting moderate-to-vigorous physical activity (MVPA), muscle-strengthening activity (MSA), and full recommendations (vs not) at follow-up 1, 2, and 3; for key predictor exercising outside (vs inside the home environment only).

Predictor	MVPA[a]		MSA[a]		Full PA[b] recommendations[a]	
	OR[c] (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Time	0.82 (0.67-1.01)	.06	0.84 (0.67-1.06)	.15	0.68 (0.52-0.89)	.005
Exercising outside (reference: inside only)	4.36 (2.87-6.63)	<.001	0.42 (0.27-0.65)	<.001	1.89 (1.10-3.23)	.02
Use of digital PA programs (reference: not)	0.95 (0.74-1.23)	.69	8.71 (6.39-11.86)	<.001	3.95 (2.84-5.50)	<.001
Time×location interaction	1.07 (0.94-1.23)	.30	1.10 (0.96-1.26)	.16	1.11 (0.93-1.32)	.26

[a]N=4439 observations, n=1769 individuals. Models fully controlled for age, gender, ethnicity, education, condition limiting PA, country, indoor space, employment, BMI, perceived risk of COVID-19, isolation status, smoking, and alcohol consumption. The Bayes factor for nonsignificant associations with digital PA program use was 0.14 (MVPA).

[b]PA: physical activity.

[c]OR: odds ratio.

Discussion

Principal Results

This study found strong associations between exercise location and digital program use with PA guideline adherence in a sample of active UK adults. Exercise location and digital program use showed different associations with MVPA, MSA, and full PA guideline adherence. Exercising inside the home environment was positively associated with MSA and exercising outside was positively associated with MVPA guideline adherence. Hence, exercising both inside and outside the home environment contributed to overall PA guideline adherence, while exercising only inside or only outside was associated with lower odds of adhering to full PA guidelines. Digital PA program use was also associated with MSA and full guideline adherence, but not with MVPA adherence. Furthermore, the results suggest that in the pandemic phases after the first initial confinements, active UK adults were less likely to exercise inside their home environment and to use digital PA programs. Users of digital PA programs were more likely to be younger, female, highly educated, have indoor space to exercise, and have a lower BMI. Most frequent exercise types inside the home environment included already owned indoor equipment, digital PA programs, one's own workout, or other types (eg, gardening and DIY), whereas MVPA and walking were the most frequently reported exercise types outside the home environment.

These results are partially consistent with expectation and previous literature. In this study, most participants who exercised outside of their home environments engaged in MVPAs (such as running or cycling). As most MVPAs require space, MSAs may be more feasible for home-based exercise. Thus, it is not surprising that exercise location was differently associated with

MVPA and MSA guideline adherence. However, previous cross-sectional research conducted during the first pandemic confinement in the United States found that exercising in one's home, garage, yard, or driveway was associated with higher MVPA [19]. The finding that digital PA program use was positively associated with PA was consistent with previous literature [16,27-29]. Although the absence of an association with MVPA differed from the findings of Parker et al [28], they reported similar relative trends in the odds for MVPA (OR 2.0), MSA (OR 3.3), and full guideline adherence (OR 2.7) in a sample including inactive participants, with the strongest association found for MSA guideline adherence. In the current study, digital PA program use was one of the most frequently reported ways to exercise inside one's home and likely has a stronger focus on MSA than MVPA due to feasibility in limited spaces. The attenuation of the ORs for both exercising inside the home environment and digital program use when controlling for each other in models predicting MSA and full guideline adherence further indicate a substantial amount of shared variance between these two predictors.

Regarding factors associated with PA digital program use in active adults, this study found that users of digital PA programs were more likely to be younger (<35 years), female, and highly educated, consistent with previous research [10,28,30,31]. Users were also more likely to have indoor space to exercise and to have a lower BMI. While digital PA programs may thus be able to target some groups at risk of insufficient PA during the pandemic (eg, women), they likely pose barriers for other disadvantaged groups who may not be able to benefit from digital solutions [43]. The pandemic has increased the already growing prepandemic health inequalities in the United Kingdom [44], and future research efforts should concentrate on how digital interventions can reach the groups most in need while addressing unintended adverse effects such as contextual, psychological, and socioeconomic access barriers [43]. Furthermore, while individuals with a condition limiting PA were more likely to exercise in their home environment, they were not more likely to use digital programs to exercise. This identifies a potential gap in targeted digital programs to address the needs of specific groups. Targeted digital interventions may be beneficial to individuals with conditions limiting PA and limited access to regular PA offers. In addition to identifying possible barriers, factors associated with digital program use may also represent differing preferences between groups. Further qualitative and quantitative research could examine the specific preferences, barriers, and facilitators associated with digital program use to help targeted intervention design.

Generally, fewer individuals adhere to MSA than to MVPA guidelines [45], and MSA has historically been neglected in guidelines and research [4]. Although this study found that active individuals who used digital PA programs and exercised inside their home environments were more likely to adhere to the MSA guidelines, the results also suggest that proportions dropped during the lifting of restrictions in the summer and reintroduction of restrictions in the autumn/winter of 2020. This drop also (partially) explained the decrease in MSA and full guideline adherence over time, which was seen in the attenuation of the effect of time when controlling for exercise location and digital program use. Health app engagement is often reported to be poorly sustained [46,47]. While the increased availability of digital PA programs and shift to home-based exercise may have presented an initial novelty, this may have become less attractive over the duration of the pandemic. It is also possible that the more strength-based activities at home were perceived as a substitute to aerobic activities rather than a complementary activity as advised in PA guidelines. Furthermore, exercising inside one's home environment may be less motivating, for example, due to the lack of socialization. Future research should investigate the potential of home-based and digital exercise for promotion of MSA and full guideline adherence. It should also be explored whether different type of programs (eg, delivered live or on-demand) can foster different engagement and adherence rates.

Limitations

This study is the first to investigate exercise locations, use of digital PA programs, and associations with PA guideline adherence, including MSA, during the COVID-19 pandemic in a longitudinal cohort of active UK adults. However, this study had some limitations.

First, all measures were self-reported. The agreement of self-reported with objectively measured PA varies substantially, and objective measures are often believed to be more accurate [48]. However, objective measures such as wearables are designed to track MVPA [49] and are hence less suitable to track home-based and strengthening activities [4]. Thus, future research should use both subjective and objective measures to account for a missing gold standard to capture both MVPA and MSA. Further, by using longitudinal data, any systematic measurement biases are corrected for as they would be expected to apply across waves. A second limitation of this study is its limited external validity due to the nonrepresentative sample, which was likely aggravated by attrition from baseline to FU3. Further, while the inclusion of inactive participants was considered methodologically problematic, their exclusion limited the generalizability of the results and may explain differences to the findings of other studies [28]. Considering the survey questions and aims in this study, and the bias of analyzing the full sample (as described in the Methods section), the adopted approach was deemed the most appropriate to answer this study's research questions. However, the present results can therefore not contribute to hypotheses about facilitators to exercise in inactive adults. Future research may seek to assess the association of availability (eg, download) of digital programs or apps with PA guideline adherence including inactive participants. Third, the data are observational and thus preclude causal conclusions. The association between exercise locations and digital program use with PA outcomes have plausible alternative explanations through third variables such as overall PA motivation, self-efficacy, and prepandemic PA levels, which were not assessed in this study. Equally, this study was not set up to investigate potential mediating mechanisms to explain this link. Future experimental intervention research should investigate causal links between digital PA program use and PA, and assess mediating mechanisms. Finally, the measurement of digital PA programs (online or app-based fitness classes/programs) did not provide a clear definition of different

type of programs (such as recorded/on-demand or live video–based programs) and did not explicitly include other forms of digital PA such as digitally conducted personal training sessions. Further research would benefit from distinguishing different types of digital solutions and investigating how they can implement different behavior change techniques [50].

Conclusions

This study found that exercising both inside and outside the home environment and the use of digital programs to exercise were associated with full WHO PA guideline adherence in active adults during the pandemic. Digital PA programs may be suitable to support home-based MSA and thus support full guideline adherence. However, usage prevalence dropped during the first 6 months of pandemic restrictions. It is recommended that future research should further investigate the role of different digital PA interventions to promote PA and program adherence, using experimental designs and representative samples.

Acknowledgments

This study was funded with support from Cancer Research UK (PRCRPG-Nov21\100002) and the SPECTRUM Consortium, funded by the UK Prevention Research Partnership (MR/S037519/1), an initiative funded by the UK Research and Innovation Councils, the Department of Health and Social Care (England), and the UK devolved administrations and leading health research charities.

Authors' Contributions

VS, DK, AH, EB, AF, and LS contributed to the concept and design of the study. VS was responsible for the data analysis and preparation and revision of the manuscript. All authors reviewed and commented on the manuscript and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Pandemic context: UK restrictions on exercising in gyms and organized sports at Health Behaviours During the COVID-19 Pandemic (HEBECO) data collection waves.
[DOCX File , 17 KB - formative_v6i6e35021_app1.docx]

Multimedia Appendix 2
Changes to the protocol.
[DOCX File , 13 KB - formative_v6i6e35021_app2.docx]

Multimedia Appendix 3
Details of measures.
[DOCX File , 18 KB - formative_v6i6e35021_app3.docx]

Multimedia Appendix 4
Unweighted baseline characteristics of the analytic sample and participants lost to follow-up.
[DOCX File , 14 KB - formative_v6i6e35021_app4.docx]

Multimedia Appendix 5
Unweighted descriptive statistics: exercise locations, use of digital programs, and meeting of WHO PA recommendations at FU1, FU2, and FU3.
[DOCX File , 14 KB - formative_v6i6e35021_app5.docx]

Multimedia Appendix 6
Unadjusted estimates: predictors of exercising inside (vs only outside), outside (vs only inside) the home environment, and of using digital PA programs (vs not) at FU1, FU2, and FU3.
[DOCX File , 15 KB - formative_v6i6e35021_app6.docx]

Multimedia Appendix 7
Complete case analysis: predictors of exercising inside (vs only outside), outside (vs only inside) the home environment, and of using digital PA programs (vs not) at FU1, FU2, and FU3.
[DOCX File , 15 KB - formative_v6i6e35021_app7.docx]

Multimedia Appendix 8
Predictors of meeting MVPA, MSA, and full recommendations (vs not) at FU1, FU2, and FU3. GLMM estimates with key predictors for unadjusted models, adjusted using exercising inside (vs outside the home environment only, Model 1) and using exercising outside (vs inside the home environment only, Model 2).
[DOCX File , 15 KB - formative_v6i6e35021_app8.docx]

Multimedia Appendix 9
Predictors of meeting MVPA, MSA, and full recommendations (vs not) at FU1, FU2, and FU3. Fully adjusted GLMM estimates with key predictor exercising inside (vs outside the home environment only).
[DOCX File , 16 KB - formative_v6i6e35021_app9.docx]

Multimedia Appendix 10
Complete case analysis: predictors of meeting MVPA, MSA, and full recommendations (vs not) at FU1, FU2, and FU3. GLMM estimates with key predictors only and using exercising inside (vs outside the home environment only).
[DOCX File , 15 KB - formative_v6i6e35021_app10.docx]

Multimedia Appendix 11
Posthoc analysis: associations of exercising inside the home environment with PA guideline adherence at FU1-FU3; results from fully adjusted binary logistic regression models.
[DOCX File , 13 KB - formative_v6i6e35021_app11.docx]

Multimedia Appendix 12
Predictors of meeting MVPA, MSA, and full recommendations (vs not) at FU1, FU2, and FU3. Unadjusted and fully adjusted GLMM estimates with key predictor exercising outside (vs inside the home environment only).
[DOCX File , 16 KB - formative_v6i6e35021_app12.docx]

Multimedia Appendix 13
Complete case analysis, predictors of meeting MVPA, MSA, and full recommendations (vs not) at FU1, FU2, and FU3. GLMM estimates with key predictors only and using exercising outside (vs inside the home environment only).
[DOCX File , 15 KB - formative_v6i6e35021_app13.docx]

References

1. Guthold R, Stevens GA, Riley LM, Bull FC. Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1·9 million participants. Lancet Glob Health 2018 Oct;6(10):e1077-e1086 [FREE Full text] [doi: 10.1016/S2214-109X(18)30357-7] [Medline: 30193830]
2. Physical activity. World Health Organization. 2020 Nov 26. URL: https://www.who.int/news-room/fact-sheets/detail/physical-activity [accessed 2022-05-28]
3. Active lives adult survey: mid-March to mid-May 2020: Coronavirus (Covid-19) Report. Sport England. 2020 Oct 22. URL: https://tinyurl.com/2p8a3vvy [accessed 2022-05-28]
4. Bennie JA, Shakespear-Druery J, De Cocker K. Muscle-strengthening exercise epidemiology: a new frontier in chronic disease prevention. Sports Med Open 2020 Aug 26;6(1):40 [FREE Full text] [doi: 10.1186/s40798-020-00271-w] [Medline: 32844333]
5. Physical inactivity: economic costs to NHS clinical commissioning groups. Public Health England. 2016 May 20. URL: https://www.gov.uk/government/publications/physical-inactivity-economic-costs-to-nhs-clinical-commissioning-groups [accessed 2022-05-28]
6. Stockwell S, Trott M, Tully M, Shin J, Barnett Y, Butler L, et al. Changes in physical activity and sedentary behaviours from before to during the COVID-19 pandemic lockdown: a systematic review. BMJ Open Sport Exerc Med 2021;7(1):e000960. [doi: 10.1136/bmjsem-2020-000960] [Medline: 34192010]
7. Maltagliati S, Rebar A, Fessler L, Forestier C, Sarrazin P, Chalabaev A, et al. Evolution of physical activity habits after a context change: The case of COVID-19 lockdown. Br J Health Psychol 2021 Nov;26(4):1135-1154 [FREE Full text] [doi: 10.1111/bjhp.12524] [Medline: 33822454]
8. Ammar A, Brach M, Trabelsi K, Chtourou H, Boukhris O, Masmoudi L, et al. Effects of COVID-19 home confinement on eating behaviour and physical activity: results of the ECLB-COVID19 International Online Survey. Nutrients 2020 May 28;12(6):1583 [FREE Full text] [doi: 10.3390/nu12061583] [Medline: 32481594]
9. Tison GH, Avram R, Kuhar P, Abreau S, Marcus GM, Pletcher MJ, et al. Worldwide effect of COVID-19 on physical activity: a descriptive study. Ann Intern Med 2020 Nov 03;173(9):767-770 [FREE Full text] [doi: 10.7326/M20-2665] [Medline: 32598162]

10. Wilke J, Mohr L, Tenforde AS, Edouard P, Fossati C, González-Gross M, et al. A pandemic within the pandemic? Physical activity levels substantially decreased in countries affected by COVID-19. Int J Environ Res Public Health 2021 Feb 24;18(5):2235 [FREE Full text] [doi: 10.3390/ijerph18052235] [Medline: 33668262]
11. Active lives adult survey: November 2019/20 Report. Sport England. 2021 Apr 29. URL: https://tinyurl.com/4r9j5r55 [accessed 2022-05-28]
12. Bu F, Bone JK, Mitchell JJ, Steptoe A, Fancourt D. Longitudinal changes in physical activity during and after the first national lockdown due to the COVID-19 pandemic in England. Sci Rep 2021 Sep 02;11(1):17723. [doi: 10.1038/s41598-021-97065-1] [Medline: 34475465]
13. McCarthy H, Potts HWW, Fisher A. Physical activity behavior before, during, and after COVID-19 restrictions: longitudinal smartphone-tracking study of adults in the United Kingdom. J Med Internet Res 2021 Feb 03;23(2):e23701 [FREE Full text] [doi: 10.2196/23701] [Medline: 33347421]
14. Meyer J, McDowell C, Lansing J, Brower C, Smith L, Tully M, et al. Changes in physical activity and sedentary behavior in response to COVID-19 and their associations with mental health in 3052 US adults. Int J Environ Res Public Health 2020 Sep 05;17(18):6469 [FREE Full text] [doi: 10.3390/ijerph17186469] [Medline: 32899495]
15. Cancello R, Soranna D, Zambra G, Zambon A, Invitti C. Determinants of the lifestyle changes during COVID-19 pandemic in the residents of Northern Italy. Int J Environ Res Public Health 2020 Aug 28;17(17):6287 [FREE Full text] [doi: 10.3390/ijerph17176287] [Medline: 32872336]
16. Constandt B, Thibaut E, De Bosscher V, Scheerder J, Ricour M, Willem A. Exercising in times of lockdown: an analysis of the impact of COVID-19 on levels and patterns of exercise among adults in Belgium. Int J Environ Res Public Health 2020 Jun 10;17(11):4144 [FREE Full text] [doi: 10.3390/ijerph17114144] [Medline: 32532013]
17. Rogers NT, Waterlow NR, Brindle H, Enria L, Eggo RM, Lees S, et al. Behavioral change towards reduced intensity physical activity is disproportionately prevalent among adults with serious health issues or self-perception of high risk during the UK COVID-19 lockdown. Front Public Health 2020;8:575091. [doi: 10.3389/fpubh.2020.575091] [Medline: 33102424]
18. Wang Y, Zhang Y, Bennell K, White DK, Wei J, Wu Z, et al. Physical distancing measures and walking activity in middle-aged and older residents in Changsha, China, during the COVID-19 epidemic period: longitudinal observational study. J Med Internet Res 2020 Oct 26;22(10):e21632 [FREE Full text] [doi: 10.2196/21632] [Medline: 33027035]
19. Dunton GF, Wang SD, Do B, Courtney J. Early effects of the COVID-19 pandemic on physical activity locations and behaviors in adults living in the United States. Prev Med Rep 2020 Dec;20:101241 [FREE Full text] [doi: 10.1016/j.pmedr.2020.101241] [Medline: 33173751]
20. Smith L, Jacob L, Butler L, Schuch F, Barnett Y, Grabovac I, et al. Prevalence and correlates of physical activity in a sample of UK adults observing social distancing during the COVID-19 pandemic. BMJ Open Sport Exerc Med 2020;6(1):e000850. [doi: 10.1136/bmjsem-2020-000850] [Medline: 34192006]
21. Ong J, Lau T, Massar S, Chong Z, Ng B, Koek D, et al. COVID-19-related mobility reduction: heterogenous effects on sleep and physical activity rhythms. Sleep 2021 Feb 12;44(2):ZSAA179 [FREE Full text] [doi: 10.1093/sleep/zsaa179] [Medline: 32918076]
22. Janssen X, Fleming L, Kirk A, Rollins L, Young D, Grealy M, et al. Changes in physical activity, sitting and sleep across the COVID-19 national lockdown period in Scotland. Int J Environ Res Public Health 2020 Dec 14;17(24):9362 [FREE Full text] [doi: 10.3390/ijerph17249362] [Medline: 33327556]
23. Mitchell J, Dicken S, Kale D, Herbec A, Beard E, Shahab L. Longitudinal changes and key determinants of meeting WHO recommended levels of physical activity during the COVID-19 pandemic in a UK-based sample: Findings from the HEBECO Study. medRxiv. 2021 Oct 03. URL: https://www.medrxiv.org/content/10.1101/2021.09.30.21264358v1 [accessed 2022-05-28]
24. Ferguson T, Curtis R, Fraysse F, Lagiseti R, Northcott C, Virgara R, et al. Annual, seasonal, cultural and vacation patterns in sleep, sedentary behaviour and physical activity: a systematic review and meta-analysis. BMC Public Health 2021 Jul 13;21(1):1384 [FREE Full text] [doi: 10.1186/s12889-021-11298-3] [Medline: 34256712]
25. Tucker P, Gilliland J. The effect of season and weather on physical activity: a systematic review. Public Health 2007 Dec;121(12):909-922. [doi: 10.1016/j.puhe.2007.04.009] [Medline: 17920646]
26. Herbec A, Schneider V, Fisher A, Kale D, Shahab L, Lally P. Correlates of and changes in aerobic physical activity and strength training before and after the onset of COVID-19 pandemic in the UK: Findings from the HEBECO study. MedRxiv. 2021 Jan 18. URL: https://www.medrxiv.org/content/10.1101/2021.01.16.21249925v1 [accessed 2022-05-28]
27. Marchant G, Bonaiuto F, Bonaiuto M, Guillet Descas E. Exercise and physical activity eHealth in COVID-19 pandemic: a cross-sectional study of effects on motivations, behavior change mechanisms, and behavior. Front Psychol 2021;12:618362. [doi: 10.3389/fpsyg.2021.618362] [Medline: 33692722]
28. Parker K, Uddin R, Ridgers ND, Brown H, Veitch J, Salmon J, et al. The use of digital platforms for adults' and adolescents' physical activity during the COVID-19 pandemic (Our Life at Home): survey study. J Med Internet Res 2021 Feb 01;23(2):e23389 [FREE Full text] [doi: 10.2196/23389] [Medline: 33481759]
29. Yang Y, Koenigstorfer J. Determinants of physical activity maintenance during the Covid-19 pandemic: a focus on fitness apps. Transl Behav Med 2020 Oct 08;10(4):835-842 [FREE Full text] [doi: 10.1093/tbm/ibaa086] [Medline: 32926160]

30. Elavsky S, Smahel D, Machackova H. Who are mobile app users from healthy lifestyle websites? Analysis of patterns of app use and user characteristics. Transl Behav Med 2017 Dec;7(4):891-901 [FREE Full text] [doi: 10.1007/s13142-017-0525-x] [Medline: 28929368]
31. Carroll JK, Moorhead A, Bond R, LeBlanc WG, Petrella RJ, Fiscella K. Who uses mobile phone health apps and does use matter? A secondary data analytics approach. J Med Internet Res 2017 Apr 19;19(4):e125 [FREE Full text] [doi: 10.2196/jmir.5604] [Medline: 28428170]
32. Ernsting C, Dombrowski SU, Oedekoven M, O Sullivan JL, Kanzler M, Kuhlmey A, et al. Using smartphones and health apps to change and manage health behaviors: a population-based survey. J Med Internet Res 2017 Apr 05;19(4):e101 [FREE Full text] [doi: 10.2196/jmir.6838] [Medline: 28381394]
33. Herbec A, Shahab L, Brown J, Kale D, Perski O, Lally P. Health Behaviors during the Covid-19 pandemic (HEBECO): A longitudinal study. Open Science Framework (OSF). 2020. URL: https://osf.io/mav3y/ [accessed 2022-05-28]
34. Schneider V, Kale D, Herbec A, Beard E, Fisher A, Shahab L. UK adults' exercise locations, use of digital programs and associations with physical activity during the COVID-19 pandemic. Open Science Framework (OSF). 2021. URL: https://osf.io/d86t7/ [accessed 2022-05-28]
35. Yore M, Ham S, Ainsworth B, Kruger J, Reis J, Kohl H, et al. Reliability and validity of the instrument used in BRFSS to assess physical activity. Med Sci Sports Exerc 2007 Aug;39(8):1267-1274. [doi: 10.1249/mss.0b013e3180618bbe] [Medline: 17762359]
36. 2015 BRFSS survey data and documentation. Centers for Disease Control and Prevention (CDC). 2017 Aug 11. URL: https://www.cdc.gov/brfss/annual_data/annual_2015.html [accessed 2022-05-28]
37. Alcohol Guidelines Review: Report from the Guidelines development group to the UK Chief Medical Officers. Department of Health. 2016 Jan. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/545739/GDG_report-Jan2016.pdf [accessed 2022-05-28]
38. Population estimates. Office for National Statistics. 2018. URL: https://www.ons.gov.uk/peoplepopulationandcommunity [accessed 2022-05-28]
39. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Soft 2015;67(1):1-48. [doi: 10.18637/jss.v067.i01]
40. Field A. Discovering statistics using IBM SPSS statistics. London, UK: SAGE Publications Ltd; 2013.
41. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B 2018 Dec 05;57(1):289-300. [doi: 10.1111/j.2517-6161.1995.tb02031.x]
42. Tattan-Birch H. Bayes factor calculator. BayesFactor.info. URL: https://harry-tattan-birch.shinyapps.io/bayes-factor-calculator/ [accessed 2022-05-28]
43. Alexander SA, Shareck M. Widening the gap? Unintended consequences of health promotion measures for young people during COVID-19 lockdown. Health Promot Int 2021 Dec 23;36(6):1783-1794 [FREE Full text] [doi: 10.1093/heapro/daab015] [Medline: 33604653]
44. Marmot M, Allen J, Goldblatt P, Herd E, Morrison J. Build Back Fairer: The COVID-19 Marmot Review. The Pandemic, Socioeconomic and Health Inequalities in England. The Health Foundation. 2020 Dec. URL: https://www.health.org.uk/publications/build-back-fairer-the-covid-19-marmot-review [accessed 2022-05-28]
45. De Cocker K, Teychenne M, White R, Bennie J. Adherence to aerobic and muscle-strengthening exercise guidelines and associations with psychological distress: A cross-sectional study of 14,050 English adults. Prev Med 2020 Oct;139:106192. [doi: 10.1016/j.ypmed.2020.106192] [Medline: 32640287]
46. Grady A, Yoong S, Sutherland R, Lee H, Nathan N, Wolfenden L. Improving the public health impact of eHealth and mHealth interventions. Aust N Z J Public Health 2018 Apr;42(2):118-119. [doi: 10.1111/1753-6405.12771] [Medline: 29384248]
47. Jee H. Review of researches on smartphone applications for physical activity promotion in healthy adults. J Exerc Rehabil 2017 Feb;13(1):3-11 [FREE Full text] [doi: 10.12965/jer.1732928.464] [Medline: 28349027]
48. Prince SA, Adamo KB, Hamel ME, Hardt J, Connor Gorber S, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. Int J Behav Nutr Phys Act 2008 Nov 06;5:56 [FREE Full text] [doi: 10.1186/1479-5868-5-56] [Medline: 18990237]
49. Walker RK, Hickey AM, Freedson PS. Advantages and limitations of wearable activity trackers: considerations for patients and clinicians. Clin J Oncol Nurs 2016 Dec 01;20(6):606-610. [doi: 10.1188/16.CJON.606-610] [Medline: 27857250]
50. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]

Abbreviations

BF: Bayes Factor
DIY: do it yourself
FU1: 1-month follow-up survey

FU2: 3-month follow-up survey
FU3: 6-month follow-up survey
GLMM: generalized linear mixed model
HEBECO: Health Behaviors During the COVID-19 Pandemic study
MSA: muscle-strengthening activities
MVPA: moderate-to-vigorous physical activity
OR: odds ratio
PA: physical activity
RQ: research question
UCL: University College London
WHO: World Health Organization

Edited by A Mavragani; submitted 19.11.21; peer-reviewed by D McCarthy, Z Jin; comments to author 17.02.22; revised version received 07.03.22; accepted 25.04.22; published 21.06.22.

Please cite as:
Schneider V, Kale D, Herbec A, Beard E, Fisher A, Shahab L
UK Adults' Exercise Locations, Use of Digital Programs, and Associations with Physical Activity During the COVID-19 Pandemic: Longitudinal Analysis of Data From the Health Behaviours During the COVID-19 Pandemic Study
JMIR Form Res 2022;6(6):e35021
URL: https://formative.jmir.org/2022/6/e35021
doi:10.2196/35021
PMID:35584123

Original Paper

Mandatory Vaccination Against COVID-19: Twitter Poll Analysis on Public Health Opinion

Valentin Ritschl[1,2*], MSc; Fabian Eibensteiner[3,4*], MD; Erika Mosor[1,2], PhD; Maisa Omara[1,2], BDS, MPH; Lisa Sperl[1,2], BSc; Faisal A Nawaz[5], MBBS; Chandragiri Siva Sai[6], MPharm; Merisa Cenanovic[7], PhD; Hari Prasad Devkota[8,9], PhD; Mojca Hribersek[4], MSc; Ronita De[10], PhD; Elisabeth Klager[4], MSc; Eva Schaden[4,11], MD; Maria Kletecka-Pulker[4,12], PhD; Sabine Völkl-Kernstock[4,13], PhD; Harald Willschke[4,12], MD, PhD; Christoph Aufricht[3], PhD; Atanas G Atanasov[4,14], PhD, PD; Tanja Stamm[1,2], PhD

[1]Section for Outcomes Research, Medical University of Vienna, Vienna, Austria

[2]Ludwig Boltzmann Institute for Arthritis and Rehabilitation, Vienna, Austria

[3]Division of Pediatric Nephrology and Gastroenterology, Department of Pediatrics and Adolescent Medicine, Comprehensive Center for Pediatrics, Medical University of Vienna, Vienna, Austria

[4]Ludwig Boltzmann Institute for Digital Health and Patient Safety, Medical University of Vienna, Vienna, Austria

[5]College of Medicine, Mohammed Bin Rashid University of Medicine and Health Sciences, Dubai, United Arab Emirates

[6]Amity Institute of Pharmacy, Amity University, Uttar Pradesh, India

[7]Sarajevo, Bosnia and Herzegovina

[8]Graduate School of Pharmaceutical Sciences, Kumamoto University, Kumamoto, Japan

[9]Program for Leading Graduate Schools, Kumamoto University, Kumamoto, Japan

[10]Indian Council of Medical Research-National Institute of Cholera and Enteric Diseases, West Bengal, India

[11]Department of Anaesthesia, Intensive Care Medicine and Pain Medicine, Medical University of Vienna, Vienna, Austria

[12]Institute for Ethics and Law in Medicine, University of Vienna, Vienna, Austria

[13]Department of Child and Adolescent Psychiatry, Medical University of Vienna, Vienna, Austria

[14]Institute of Genetics and Animal Biotechnology of the Polish Academy of Sciences, Jastrzebiec, Poland

[*]these authors contributed equally

Corresponding Author:
Tanja Stamm, PhD
Section for Outcomes Research
Medical University of Vienna
Spitalgasse 23
Vienna, 1190
Austria
Phone: 43 140400 ext 16370
Email: tanja.stamm@meduniwien.ac.at

Abstract

Background: On January 30, 2020, the World Health Organization Emergency Committee declared the rapid worldwide spread of COVID-19 a global health emergency. By December 2020, the safety and efficacy of the first COVID-19 vaccines had been demonstrated. However, international vaccination coverage rates have remained below expectations (in Europe at the time of manuscript submission). Controversial mandatory vaccination is currently being discussed and has already been introduced in some countries (Austria, Greece, and Italy). We used the Twitter survey system as a viable method to quickly and comprehensively gather international public health insights on mandatory vaccination against COVID-19.

Objective: The purpose of this study was to better understand the public's perception of mandatory COVID-19 vaccination in real time using Twitter polls.

Methods: Two Twitter polls were developed (in the English language) to seek the public's opinion on the possibility of mandatory vaccination. The polls were pinned to the Digital Health and Patient Safety Platform's (based in Vienna, Austria) Twitter timeline for 1 week in mid-November 2021, 3 days after the official public announcement of mandatory COVID-19 vaccination in Austria. Twitter users were asked to participate and retweet the polls to reach the largest possible audience.

Results: Our Twitter polls revealed two extremes on the topic of mandatory vaccination against COVID-19. Almost half of the 2545 respondents (n=1246, 49%) favor mandatory vaccination, at least in certain areas. This attitude contrasts with the 45.7% (n=1162) who categorically reject mandatory vaccination. Over one-quarter (n=621, 26.3%) of participating Twitter users said they would never get vaccinated, as reflected by the current Western European and North American vaccination coverage rate. Concatenating interpretation of these two polls should be done cautiously as participating populations might substantially differ.

Conclusions: Mandatory vaccination against COVID-19 (in at least certain areas) is favored by less than 50%, whereas it is opposed by almost half of the surveyed Twitter users. Since (social) media strongly influences public perceptions and views, and social media discussions and surveys are specifically susceptible to the "echo chamber effect," the results should be interpreted as a momentary snapshot. Therefore, the results of this study need to be complemented by long-term surveys to maintain their validity.

(JMIR Form Res 2022;6(6):e35754) doi:10.2196/35754

KEYWORDS

COVID-19; SARS-CoV-2; vaccine; vaccination; Twitter; survey; mandatory vaccination; vaccination hesitancy; coronavirus; hesitancy; social media; questionnaire; mandatory; support; poll; opinion; public health; perception

Introduction

Many mitigation measurements, mostly nonpharmaceutical interventions, have been undertaken on local, national, and international levels to reduce the transmission of COVID-19 since the beginning of this pandemic [1-6]. As Sridhar and Gurdasani [7] discussed in January 2021, immunity can be boosted safely through vaccination in many infectious diseases, while achievement of herd immunity through SARS-CoV-2 infection is not a strategy worth considering due to little guarantee of success while putting a high toll on morbidity and mortality. Therefore, the return to prepandemic normality may rely on the success of vaccine-induced immunity to prevent severe disease and limit dissemination [8]. Although the first COVID-19 vaccines were quickly proven safe and efficacious, and were approved by regulatory authorities in December 2020 [9-12], global vaccination coverage has not been achieved for several reasons. Vaccination coverage rates largely depend on a country's wealth and other factors that influence the vaccination behavior of a country's citizens. As a result, SARS-CoV-2 variants continue to emerge, triggering disease episodes and slowing or even reversing the reopening of societies and economies [13].

Widespread public acceptance of vaccines continues to be a challenging endeavor requiring accountancy of complex socioeconomic factors on the level of international policy makers, national and local public health officials, and professional and community organizations [14]. A large study from four metropolitan areas of the United States found more than 20% of participants reluctant to vaccinate. Participants expressed concerns on efficacy and safety while also questioning the severity of a COVID-19 infection [15]. In a Canadian study, participants who did not plan to get vaccinated were also less likely to retain mitigation measures such as wearing face masks and practicing physical distancing [16].

Outreach to the public providing necessary information regarding COVID-19 has been achieved over several communication channels, such as traditional media and different social media platforms [17]. Public health implications of social media platforms such as Twitter have been studied before and with increasing intensity during the COVID-19 pandemic. Examples include public perception of antibiotic use and misuse, human papillomavirus vaccination on Twitter and analysis of boosted vaccination hesitancy, and the re-emergence of measles in the United States after its elimination [18-20]. During the COVID-19 pandemic, substantial effort has also been drawn to study symptoms for COVID-19 screening, dissemination of medical information and misinformation, the emergence of conspiracy theories, and discussions and emotions associated with COVID-19 on Twitter [21-25]. Longitudinal sentiment analysis of Twitter discussions around COVID-19 revealed a peak of percentages of tweets expressing fear in mid-March 2020 after the initial declaration of the pandemic, with the lowest point in early November 2020 when the first COVID-19 vaccines were announced. With the increasing perspective of promising vaccine results, the percentage of tweets expressing trust increased while fear declined [26]. With Twitter hosting about 353 million monthly active users and incorporating an inbuilt anonymous polling tool, it allows for potential insights into pressing public health topics on an international level with real-time feedback [27,28]. In a previous Twitter poll study on the public's perceptions of the currently available COVID-19 vaccines, Eibensteiner et al [29] detected a high willingness to get vaccinated despite high levels of uncertainty regarding the available vaccine's safety in February 2021. A later published study analyzing 4 million tweets since the beginning of the pandemic added that Twitter bots or political activists partly generated vaccine opposition content.

In contrast, positive content on COVID-19 vaccination was produced mainly by well-known individuals and organizations [30]. A recent study by Germani and Biller-Andorno [31] also shows that those against vaccination increasingly participate in discussions on Twitter and disseminate their content from a pool of strong influencers such as political activists, authors, or artists. Donald Trump, a previous president of the United States, was the most influential disseminator of antivaccination content on Twitter (before his account was suspended). At this point, it should be emphasized that, of course, not only social media such as Twitter but also other factors can lead to vaccination hesitancy. Truong et al [32] mentioned in their review, for example, demographic factors (ethnicity, age, gender, pregnancy, education, and employment), personal responsibility

and risk perception, trust in health authorities, and the (perceived) safety and efficacy of a new vaccine, as well as a lack of information or incorrect information about vaccines.

Mandatory vaccination against COVID-19 for the public or health care workers has been a recent focus of attention in many European countries, including Austria, Germany, and the United Kingdom [33-35]. In Austria (general public 18 years or older) and Germany (employees in hospitals and care facilities), mandatory vaccination is scheduled for spring 2022. Recently, Italy approved mandated vaccines for everyone over 50 years of age and Greece for people older than 60 years [34-36]. A survey in Germany on 20,000 households in early summer 2020 revealed that about 50% of Germany's residents would favor mandatory vaccination [37].

With the recent announcement of mandatory COVID-19 vaccination in some western democracies (eg, Austria or Germany), our study aims to survey the public's attitude on this matter. As has already been shown before, the fast-paced dynamic of this pandemic requires online survey tools to gain immediate large-scale international public health insights [23,26,29,38]. Therefore, we used the Twitter polling tool to rapidly collect and analyze the public's opinion on mandatory COVID-19 vaccination to understand endorsement and refusal, possibly aiding policy makers on this highly relevant and timely topic. This study aimed to better understand public perceptions of mandatory COVID-19 vaccination in real time using Twitter polls.

Methods

Overview

To meet the objective of this study and better understand public perceptions of mandatory COVID-19 vaccination, we conducted two Twitter polls. For this purpose, we used the Twitter account of the Digital Health and Patient Safety Platform (DHPSP; Twitter handle @DHPSP) [39]. The DHPSP was founded by the Ludwig Boltzmann Institute for Digital Health and Patient Safety, established in Austria in 2019 [29].

For this study, we distributed two Twitter polls online via the Twitter account @DHPSP between November 22 and 29, 2021. The polls were developed within the project expert team and evaluated in multiple rounds to ensure the best possible readability and comprehensibility. Poll 1 addressed whether participants had already been vaccinated ("Have you been vaccinated against COVID-19?"), whereas poll 2 asked about opinions about a possible mandatory vaccination for COVID-19 ("Do you support mandatory vaccination against COVID-19?"). Both polls were linked (poll 2 was posted as a comment under poll 1) and pinned to the top of the DHPSP Twitter timeline during the poll period. Pinning a tweet permanently places it at the top of a Twitter user's account so that any new visitors will see this tweet at the top of the visited user's timeline. The poll questions, including relevant hashtags for categorization, are limited to 280 characters on Twitter. Twitter allows up to four responses with a limit of 25 characters including spaces for each poll. Therefore, both polls had four responses ranging from complete agreement ("Yes, twice or more" and "Yes, definitely") to complete disagreement ("I never will" and "I am clearly against it") in the manner of a four-point response scale. Both surveys were categorized with the following hashtags to increase visibility and facilitate analysis: #MandatoryVaccination, #COVID19vaccines, and #DHPSP. Figure 1 shows the detailed structure of the two polls as they were distributed on Twitter.

Once the polls were launched, the first accounts to see them in their Twitter timelines were DHPSP Twitter followers. Twitter polls are anonymous and do not allow respondents' characteristics (eg, gender) to be assessed. Therefore, to obtain the characteristics of the audience that were first exposed to the polls, we attempted to analyze the characteristics of @DHPSP's followers via the online tool Followerwonk [40] on December 1, 2021. In total, the Twitter account @DHPSP had 943 followers at the time of the analysis. Of these, 206 (21.8%) were male, 133 (14.1%) were female, and 604 (64.1%) did not indicate their gender on Twitter. Overall, 137 (14.5%) DHPSP followers had more than 5000 followers on their own, 320 (33.9%) had between 500 and 5000 followers, and 486 (51.5%) had less than 500 followers. The geographical distribution of the @DHPSP follower network can be seen in Figure 2. In addition, the DHPSP network includes 225 people on the mailing list, 306 people on LinkedIn, and 1757 Facebook followers.

Figure 1. Structure of the two Twitter polls. DHPSP: Digital Health and Patient Safety Platform.

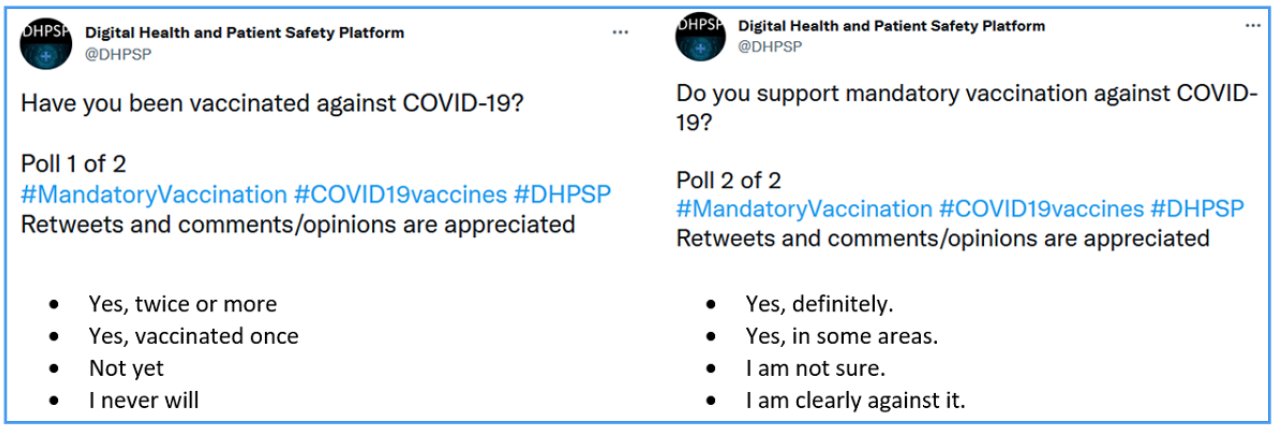


Figure 2. Main locations of the Digital Health and Patient Safety Platform's (DHPSP) Twitter followers (note: these data cover only the fraction of the DHPSP's followers who indicated their location in their account information on Twitter).

The text of the polls asked for retweets and discussion ("Retweets and comments/opinions are appreciated"; Figure 1), and with each new retweet, the polls gained an additional audience (consisting of the followers of the retweeting accounts). To gain extra visibility, members and subscribers to the DHPSP platform email list [39] were asked to support the polls by voting, retweeting, and sharing them through various additional networking approaches via emails or direct messages on social media. Various social media accounts of DHPSP members were also used to post hyperlinks to the Twitter polls. In addition, information about the polls was disseminated through DHPSP's Facebook [41] and LinkedIn accounts [42].

To characterize the user population that retweeted the polls under study, we conducted a hashtag analysis using the Symplur Signals online tool [43]. We analyzed (in terms of the number of retweets, users, locations, and languages) all tweets that contained the unique combination of the hashtags #MandatoryVaccination, #COVID19vaccines, and #DHPSP at the end of the poll period on November 29, 2021. To ensure accuracy and limit interference from other Twitter discussions on this topic, a Twitter search was conducted before the start of the surveys on November 21, 2021, which confirmed that this hashtag combination had never been used before. Symplur Signals was also used for sentiment analysis of all tweets containing this unique combination of hashtags. For additional analysis of the gender and age distribution of direct retweets, the tool Tweepsmap was applied using its "Tweet Reach" feature (quote tweets not included due to the specifics of the used tool).

Ethical Considerations

No ethical approval was required for this study as it does not fall within the scope of the Austrian Medical Ethics Act. Individual votes, retweets, and comments of any kind were anonymized using Symplur Signals. All data presented in this manuscript are anonymous. Thus, the data collected do not fall within the scope of the General Data Protection Regulation [44]. As follow-up information, the voting counts are immediately revealed to users as soon as they vote through their Twitter account. The parameters analyzed beyond the survey results, such as the number of followers and retweets, are based on online publicly available data.

Results

Both Twitter polls were pinned to the timeline of the DHPSP's Twitter account (Twitter handle @DHPSP; Table 1) for 7 days, beginning on November 22, 2021. The gender and age distribution of the Twitter accounts that directly retweeted both polls (poll 1, n=178; poll 2, n=189) are depicted in Table 2. Sentiment analysis of all tweets featuring the polls demonstrated 45% (32 primary tweets with a sentiment score range from –0.5894 to –0.062) negative sentiment and 55% (50 primary tweets with a sentiment score range from 0.9338 to 0.7536) positive sentiment of the analyzed tweets (Figure 3).

Poll 1 ("Have you been vaccinated against COVID-19?") received a total of 2365 votes (199,902 views), whereas poll 2 ("Do you support mandatory vaccination against COVID-19?") received a total of 2545 votes (200,939 views). Upon analysis of the polls' retweets that contained the unique combination of the hashtags #MandatoryVaccination, #COVID19vaccines, and #DHPSP, a total of 2073 tweets from 442 users (one retweet: n=272, 61.5%; two retweets: n=100 users, 22.6%; three or more retweets: n=70, 15.8%) were identified. The polls, including all retweets, summed up to a total of 32,594,283 views on Twitter. The top locations of Twitter users retweeting the polls were the United States (n=59, 6.3%), Canada (n=41, 4.5%), and the United Kingdom (n=17, 1.5%). However, most of the users did not indicate their location. A summary of these details is given in Table 3.

Of the Twitter users who responded to poll 1 ("Have you been vaccinated against COVID-19?"), 63.4% (1499/2365) agreed with "Yes, twice or more"; therefore, almost two-thirds of users who answered this question reported to be fully vaccinated. More than one-quarter (621/2365, 26.3%) of Twitter users

expressed that they will never get vaccinated against COVID-19 (they voted "I never will"). Together, these two groups represent the extremes regarding vaccination and represent 89.6% (2120/2365) of all answers to this question. Of the 2365 respondents, 111 (4.7%) reported that they have not yet been vaccinated but do not rule out the possibility of getting vaccinated ("Not yet."). The remaining 5.7% (n=134) of Twitter users had been vaccinated but have not yet received a second vaccine dose ("Yes, vaccinated once").

Poll 2 examined Twitter users' attitudes toward possible mandatory vaccination ("Do you support mandatory vaccination against COVID-19?"). In this poll, 40.2% (1022/2545) of all respondents indicated ("Yes, definitely") that they would support mandatory vaccination against COVID-19. In addition, 8.8% (224/2545) indicated that they would support mandatory vaccination in certain areas (eg, for certain professions; "Yes, in some areas"). In contrast to the vaccine supporters, almost half of the respondents (1162/2545, 45.7%) are strictly against mandatory vaccination ("I am clearly against it"). A small percentage of 5.4% (137/2545) of Twitter users stated that they do not yet have an opinion on mandatory vaccination against COVID-19 ("I am not sure"). A detailed summary of the responses to both polls is given in Figure 4.

Table 1. Digital Health and Patient Safety Platform's Twitter follower characteristics.

	Followers (n=943), n (%)
Gender	
Male	206 (21.8)
Female	133 (14.1)
Not stated	604 (64.1)
Follower counts	
<500	486 (51.5)
500-5000	320 (33.9)
>5000	137 (14.5)
Account ages (years)	
<1	86 (9.1)
1-5	325 (34.5)
>5	532 (56.4)
Languages	
English	586 (62.1)
Spanish	46 (4.9)
Other	301 (31.9)

Table 2. Gender and age distribution of Twitter accounts that retweeted the polls.

	Twitter accounts, n (%)
Poll 1 (n=178)	
Distribution by gender poll 1: direct retweets	
Male	102 (57.3)
Female	61 (34.4)
Businesses/groups	15 (8.3)
Distribution by age poll 1: direct retweets	
10-23 years	20 (11.4)
24-64 years	142 (80.0)
≥65 years	16 (8.6)
Poll 2 (n=189)	
Distribution by gender poll 2: direct retweets	
Male	112 (59.6)
Female	69 (36.4)
Businesses/groups	8 (4)
Distribution by age poll 2: direct retweets	
10-23 years	28 (14.8)
24-64 years	154 (81.5)
≥65 years	7 (3.7)

Figure 3. Sentiment analysis of the poll's retweets containing the unique combination of the following hashtags: #MandatoryVaccination, #COVID19vaccines, and #DHPSP (Digital Health and Patient Safety Platform). The upper panel indicates sentiment over time and the overall sentiment scores (45% negative and 55% positive), and the lower panel displays sentiment word frequencies.

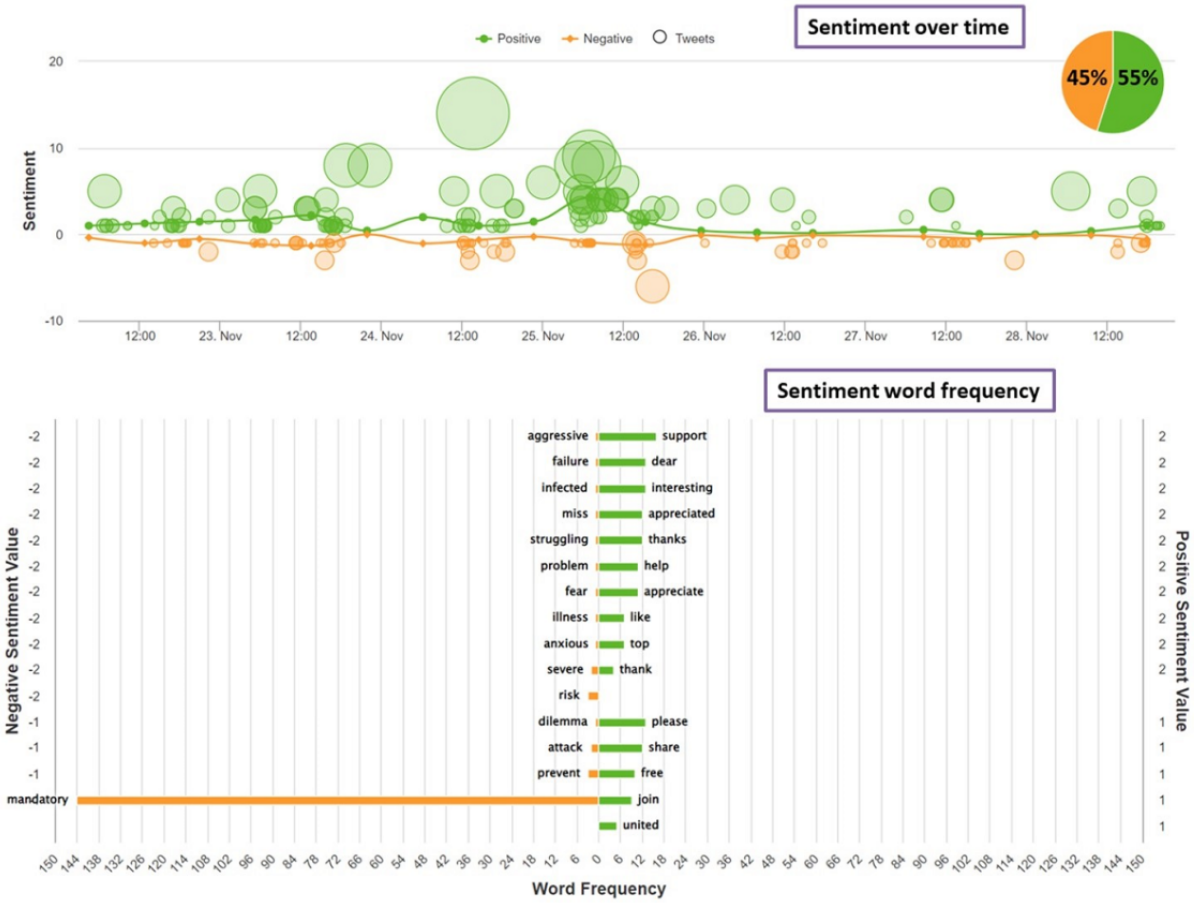


Table 3. Analysis of the poll's retweets containing the unique combination of the following hashtags: #MandatoryVaccination, #COVID19vaccines, and #DHPSP (Digital Health and Patient Safety Platform).

	Twitter users (n=943), n (%)
Top locations of Twitter users[a]	229 (24.3)
United States	59 (6.3)
Canada	41 (4.5)
United Kingdom	17 (1.5)
Number of users that retweeted[b]	442 (46.9)
One retweet	272 (61.5)
Two retweets[c]	100 (22.6)
Three or more retweets[c]	70 (15.8)
Top languages[d]	
English	1969 (95.0)
Other languages	104 (5.0)

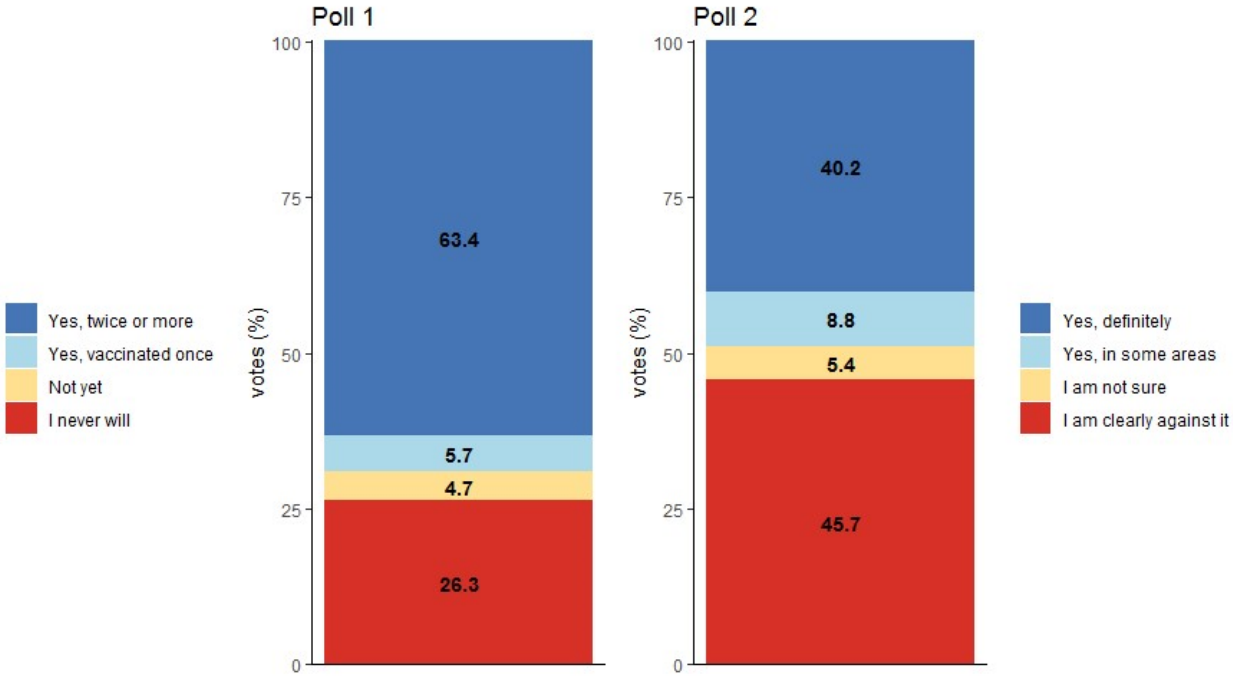
[a]Determined based on data derived just from the users who indicated their location in their account information on Twitter. While interpreting the data, the readers should be aware that 75.7% of the 943 users did not provide location information on their profiles.

[b]The total number of retweets was 2073.

[c]Including "regular" retweets, retweets with comments, and "quote retweets" (whereby a hyperlink to the original tweet is inserted in the newly composed tweet).

[d]The most used languages are indicated. All other tweet languages accounted for less than 0.5% each.

Figure 4. Twitter users' answers to poll 1 ("Have you been vaccinated against COVID-19?"; respondents n=2365) and poll 2 ("Do you support mandatory vaccination against COVID-19?"; respondents n=2545).



Discussion

Our Twitter polls showed two extremes on the issue of compulsory vaccination against COVID-19. Almost half of the respondents favor compulsory vaccination, at least for certain professional groups. However, this is in contrast to nearly as many people who categorically reject compulsory vaccination, indicating that a proportion of vaccinated people voted against mandating vaccination. In line with recent works building the methodological basis and outlining the possible benefits of Twitter polling to gain quick insights into the public's attitudes on timely matters [28,29], we aimed to explore the public's opinion on mandatory COVID-19 vaccination by using Twitter polls. This topic is currently of high interest as recently mandatory COVID-19 vaccination has been scheduled for spring 2022 in Austria and Germany [34,35]. This timely and important Twitter survey reveals that mandatory vaccination against COVID-19 (in at least certain areas) is supported by less than 50%, whereas it is opposed by almost half of the surveyed Twitter users.

We used an established Twitter network (the DHPSP's Twitter account) to pin our Twitter polls to generate high outreach and a high number of respondents in this study. This work complements our initial Twitter survey on the perceived safety of the available COVID-19 vaccines and participants' confidence or hesitancy to get vaccinated [29]. In February 2021, 83% of Twitter users participating in polls posted back then stated that they would definitely get vaccinated against COVID-19. About 70% of participants indicated that they received at least their first dose in the current poll. The percentage of participants expressing their reluctance to get vaccinated increased from 8% to 26%. This data is in line with the current vaccination rates in Western European countries (United Kingdom, Germany, Austria, France, Spain, and Italy) and the United States, with 71% to 83% of people being at least partly vaccinated [45], and above the world average of 55% [45]. A direct comparison of these two Twitter polls can be carefully made, considering that the initial Twitter network was the same, even though participating users might substantially differ. Interpretability is further limited by the DHPSP's follower base consisting mainly of younger educated individuals with scientific backgrounds, mainly due to the selective science-based content that attracted such followers and their interest in science. This concern is in line with several studies associating COVID-19 vaccine hesitancy with lower educational levels [46-49]. Comparability of other studies with our study is not straightforward, as the logic and infrastructure of Twitter surveys make the retrospective characterization of the sample difficult. Available information about Twitter users is limited (eg, in terms of location and language). Despite the composition of the original sample of "younger, educated people," we reached a total of 26.3% (621/2365) of convinced nonvaxxers in our survey, which is in line with the results reported in other studies, such as the one by Hacquin and colleagues [49]. We explain this phenomenon by the fact that while the original sample was indeed enriched with younger, (better formally) educated individuals, each new retweet immediately adds the followers of each new retweeting account as an audience for the surveys. In this way, we seem to have achieved an equilibrium consistent with opinion patterns among Twitter users, which may explain why the proportion of nonvaxxers in our study was similar to that observed in previous work.

Our findings agree with the previously reported percentage of 26% of participants stating that they will never get vaccinated against COVID-19, with about 5% remaining undecided. In the current literature, COVID-19 vaccine hesitancy is further evident among university students (14%), medical students (23%), and health care workers (28%) [50-52]. A large-scale survey in

Ireland and the United Kingdom segmented the previously published high percentages of COVID-19 vaccination hesitancy/refusal (eg, 33% in the United States) to 25% to 26% vaccination hesitancy and 6% to 9% resistance [53]. Although we aimed to formulate our poll questions and answers clearly and discriminatively (eg, "I never will"), further interpretation of these results due to missing demographics and other indicators of socioeconomic status, health literacy, and political and religious views was not possible. In addition, the study published by Murphy et al [53] was conducted in Spring 2020, at a time when no COVID-19 vaccines were yet available. Vaccination refusal may not only affect the course of this pandemic but has also been the focus of discussion among the prevention of other vaccine-preventable and potentially deadly diseases such as measles [54].

Our Twitter polls revealed high rates of being vaccinated against COVID-19 at least once, while 40% of participants supported mandatory vaccination for the public. Such concatenating interpretation of these two polls needs to be done cautiously, as the participating populations might substantially differ between both polls. Mandatory vaccination against COVID-19 is currently under rigorous discussion in many European countries and among health and home care workers [33,55,56]. Austria is currently the first western democracy, followed by Germany, to officially announce mandatory vaccination against COVID-19 for the public [34,35]. This would not be the first time mandatory vaccination has been enforced. For example, in England and Imperial Germany, between 1874 and 1975, vaccination against smallpox was compulsory, resulting in substantially reduced mortality rates [57]. However, ethical, medical, and philosophical reasons supporting and opposing mandatory vaccination involve complex socioeconomic and psychological perspectives [58-60] that are out of this paper's scope and will not be further discussed.

Several studies have assessed the public's opinion on mandatory COVID-19 vaccination in German, French, Greek, Austrian, American, Pakistani, and Italian individuals. A broad range favored mandatory COVID-19 vaccination from 17% to 74% of survey participants. Five of these studies were conducted before COVID-19 vaccines were available and before the official public announcement of mandatory vaccination against COVID-19 for the public in Austria [37,61-66].

Therefore, this is the first study analyzing the public's attitude on mandatory COVID-19 vaccination immediately following (3 days) the official public announcement of the Austrian government, which is the first western democracy to mandate vaccination for its entire population [34]. This timely analysis on an international scale provides valuable insights into changes in the public's attitude and general beliefs toward COVID-19 vaccine mandates that may aid policy makers in strategizing similar paths to Austria.

The strengths of this study lay in the rapid and timely assessment on an international scale with clear and concise information on the publics' attitude. As previously discussed in our prequel study, restrictions of word counts in Twitter polling might serve as a strength for concise and well-formulated surveys, possibly aiding a higher number of respondents in comparison to traditional surveys. On the other hand, it substantially limits the interpretability of results, as no additional information may be retrieved, and the formulation of more complex questions and clarification statements is hampered [29]. For example, it could be that the wording of the second answer of the second poll, "yes, in some areas" I am in favor of compulsory vaccination, may be understood by some respondents as "region," "district," or another geographical unit. In the original conception, however, the term "area" was meant rather as "branch" or "profession." A misunderstanding cannot be ruled out and must be considered when interpreting the results. Distribution of the Twitter polls via a pre-existing network might also influence sample selection and cause bias further challenged by a lack of baseline characteristics of the survey participants. This is potentiated by the "echo chamber effect," exposing social media users to curated content most likely aligning with their pre-existing beliefs based on their previous social media behavior [67,68]. This also includes the previous observation that individuals exposed to negative opinions on human papillomavirus vaccination were more likely to share these on Twitter, in contrast to those exposed to neutral or positive opinions [69]. An explicit limitation also lies in the extremely limited concatenating interpretation of the different Twitter polls, as the participating user populations might substantially differ between the two polls, and comparability cannot be achieved due to the polls' anonymity. Thus, a direct transferability of the results to individual countries is impossible or only possible to a limited extent. In addition, poll manipulation by exploiting multiple users due to the polls' anonymity needs to be kept in mind upon interpreting these results, as already mentioned by Vidal-Alaball et al [28]. We used Symplur Signals for sentiment analysis of all tweets. This automated text-mining tool helped us get an impression of the sentiment. Still, it has to be mentioned that this is not equal to classical qualitative data analysis.

When interpreting the data, it is important to remember that the two surveys involved self-selected users of a popular social media platform who self-reported their COVID-19 vaccination status and their opinion on introducing such mandatory vaccination. Objective data, for example, on vaccination status, could not be collected in this study.

In summary, mandatory vaccination against COVID-19 is supported by less than 50% of Twitter users and opposed by almost half of the Twitter users surveyed in this study. Refusal rates of COVID-19 vaccination are prevalent among 26% of surveyed Twitter users. These findings are reflected by the current vaccination coverage rates and align with the existing literature. Public perceptions and views on health issues are heavily influenced by social media, being specifically susceptible to the "echo chamber effect," underscoring the importance of using social media surveys to understand the public's views on health in real time to inform public health messages and communications efforts.

Acknowledgments

This research received no external funding.

Data Availability

Raw data are available upon reasonable request.

Authors' Contributions

VR, FE, AGA, and TS conceptualized and planned the study, including the preparation of the polls. VR, FE, CA, AGA, and TS wrote the first draft of the manuscript. CA, AGA, and TS reviewed the whole process. All other authors suggested and agreed upon the research questions, disseminated the polls via their networks, read the report prior to the manuscript, discussed the results, and contributed to the text. All authors approved the final version of the manuscript.

Conflicts of Interest

TS reports grants and personal fees from AbbVie, grants and personal fees from Roche, personal fees from Sanofi, personal fees from Takeda, and personal fees from Novartis, outside the submitted work. The other authors declare no conflicts of interest.

References

1. Anderson M, Mckee M, Mossialos E. Covid-19 exposes weaknesses in European response to outbreaks. BMJ 2020 Mar 18;368:m1075. [doi: 10.1136/bmj.m1075] [Medline: 32188590]
2. Anderson RM, Heesterbeek H, Klinkenberg D, Hollingsworth TD. How will country-based mitigation measures influence the course of the COVID-19 epidemic? Lancet 2020 Mar 21;395(10228):931-934 [FREE Full text] [doi: 10.1016/S0140-6736(20)30567-5] [Medline: 32164834]
3. Chu D, Akl EA, Duda S, Solo K, Yaacoub S, Schünemann HJ, COVID-19 Systematic Urgent Review Group Effort (SURGE) study authors. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. Lancet 2020 Jun 27;395(10242):1973-1987 [FREE Full text] [doi: 10.1016/S0140-6736(20)31142-9] [Medline: 32497510]
4. Cowling B, Ali S, Ng T, Tsang T, Li J, Fong M, et al. Impact assessment of non-pharmaceutical interventions against coronavirus disease 2019 and influenza in Hong Kong: an observational study. Lancet Public Health 2020 May;5(5):e279-e288 [FREE Full text] [doi: 10.1016/S2468-2667(20)30090-6] [Medline: 32311320]
5. Eibensteiner F, Ritschl V, Ariceta G, Jankauskiene A, Klaus G, Paglialonga F, European Pediatric Dialysis Working Group. Rapid response in the COVID-19 pandemic: a Delphi study from the European Pediatric Dialysis Working Group. Pediatr Nephrol 2020 Sep;35(9):1669-1678 [FREE Full text] [doi: 10.1007/s00467-020-04584-6] [Medline: 32418146]
6. Pan A, Liu L, Wang C, Guo H, Hao X, Wang Q, et al. Association of public health interventions with the epidemiology of the COVID-19 outbreak in Wuhan, China. JAMA 2020 May 19;323(19):1915-1923 [FREE Full text] [doi: 10.1001/jama.2020.6130] [Medline: 32275295]
7. Sridhar D, Gurdasani D. Herd immunity by infection is not an option. Science 2021 Jan 15;371(6526):230-231. [doi: 10.1126/science.abf7921] [Medline: 33446540]
8. Rubin EJ, Longo DL. SARS-CoV-2 vaccination - an ounce (actually, much less) of prevention. N Engl J Med 2020 Dec 31;383(27):2677-2678 [FREE Full text] [doi: 10.1056/NEJMe2034717] [Medline: 33301245]
9. Baden LR, El Sahly HM, Essink B, Kotloff K, Frey S, Novak R, COVE Study Group. Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. N Engl J Med 2021 Feb 04;384(5):403-416 [FREE Full text] [doi: 10.1056/NEJMoa2035389] [Medline: 33378609]
10. Logunov D, Dolzhikova I, Shcheblyakov D, Tukhvatulin A, Zubkova O, Dzharullaeva A, Gam-COVID-Vac Vaccine Trial Group. Safety and efficacy of an rAd26 and rAd5 vector-based heterologous prime-boost COVID-19 vaccine: an interim analysis of a randomised controlled phase 3 trial in Russia. Lancet 2021 Feb 20;397(10275):671-681 [FREE Full text] [doi: 10.1016/S0140-6736(21)00234-8] [Medline: 33545094]
11. Polack FP, Thomas SJ, Kitchin N, Absalon J, Gurtman A, Lockhart S, C4591001 Clinical Trial Group. Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. N Engl J Med 2020 Dec 31;383(27):2603-2615 [FREE Full text] [doi: 10.1056/NEJMoa2034577] [Medline: 33301246]
12. Voysey M, Clemens S, Madhi S, Weckx L, Folegatti P, Aley P, Oxford COVID Vaccine Trial Group. Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK. Lancet 2021 Jan 09;397(10269):99-111 [FREE Full text] [doi: 10.1016/S0140-6736(20)32661-1] [Medline: 33306989]
13. Tregoning JS, Flight KE, Higham SL, Wang Z, Pierce BF. Progress of the COVID-19 vaccine effort: viruses, vaccines and variants versus efficacy, effectiveness and escape. Nat Rev Immunol 2021 Oct;21(10):626-636 [FREE Full text] [doi: 10.1038/s41577-021-00592-1] [Medline: 34373623]

14. Schoch-Spana M, Brunson EK, Long R, Ruth A, Ravi SJ, Trotochaud M, et al. The public's role in COVID-19 vaccination: human-centered recommendations to enhance pandemic vaccine awareness, access, and acceptance in the United States. Vaccine 2021 Sep 24;39(40):6004-6012 [FREE Full text] [doi: 10.1016/j.vaccine.2020.10.059] [Medline: 33160755]
15. El-Mohandes A, White TM, Wyka K, Rauh L, Rabin K, Kimball SH, et al. COVID-19 vaccine acceptance among adults in four major US metropolitan areas and nationwide. Sci Rep 2021 Nov 04;11(1):21844. [doi: 10.1038/s41598-021-00794-6] [Medline: 34737319]
16. Muhajarine N, Adeyinka DA, McCutcheon J, Green KL, Fahlman M, Kallio N. COVID-19 vaccine hesitancy and refusal and associated factors in an adult population in Saskatchewan, Canada: Evidence from predictive modelling. PLoS One 2021;16(11):e0259513 [FREE Full text] [doi: 10.1371/journal.pone.0259513] [Medline: 34767603]
17. Su Y. It doesn't take a village to fall for misinformation: social media use, discussion heterogeneity preference, worry of the virus, faith in scientists, and COVID-19-related misinformation beliefs. Telematics Inform 2021 May;58:101547. [doi: 10.1016/j.tele.2020.101547]
18. Hotez PJ, Nuzhath T, Colwell B. Combating vaccine hesitancy and other 21st century social determinants in the global fight against measles. Curr Opin Virol 2020 Apr;41:1-7. [doi: 10.1016/j.coviro.2020.01.001] [Medline: 32113136]
19. Kendra RL, Karki S, Eickholt JL, Gandy L. Characterizing the discussion of antibiotics in the Twittersphere: what is the bigger picture? J Med Internet Res 2015 Jun 19;17(6):e154 [FREE Full text] [doi: 10.2196/jmir.4220] [Medline: 26091775]
20. Massey PM, Leader A, Yom-Tov E, Budenz A, Fisher K, Klassen AC. Applying multiple data collection tools to quantify human papillomavirus vaccine communication on Twitter. J Med Internet Res 2016 Dec 05;18(12):e318 [FREE Full text] [doi: 10.2196/jmir.6670] [Medline: 27919863]
21. Ahmed W, Vidal-Alaball J, Downing J, López Seguí F. COVID-19 and the 5G conspiracy theory: social network analysis of Twitter data. J Med Internet Res 2020 May 06;22(5):e19458 [FREE Full text] [doi: 10.2196/19458] [Medline: 32352383]
22. Guo J, Radloff CL, Wawrzynski SE, Cloyes KG. Mining twitter to explore the emergence of COVID-19 symptoms. Public Health Nurs 2020 Nov;37(6):934-940 [FREE Full text] [doi: 10.1111/phn.12809] [Medline: 32937679]
23. Krittanawong C, Narasimhan B, Virk HUH, Narasimhan H, Hahn J, Wang Z, et al. Misinformation dissemination in Twitter in the COVID-19 era. Am J Med 2020 Dec;133(12):1367-1369 [FREE Full text] [doi: 10.1016/j.amjmed.2020.07.012] [Medline: 32805227]
24. Rosenberg H, Syed S, Rezaie S. The Twitter pandemic: the critical role of Twitter in the dissemination of medical information and misinformation during the COVID-19 pandemic. CJEM 2020 Jul;22(4):418-421 [FREE Full text] [doi: 10.1017/cem.2020.361] [Medline: 32248871]
25. Xue J, Chen J, Hu R, Chen C, Zheng C, Su Y, et al. Twitter discussions and emotions about the COVID-19 pandemic: machine learning approach. J Med Internet Res 2020 Nov 25;22(11):e20550 [FREE Full text] [doi: 10.2196/20550] [Medline: 33119535]
26. Lyu JC, Han EL, Luli GK. COVID-19 vaccine-related discussion on Twitter: topic modeling and sentiment analysis. J Med Internet Res 2021 Jun 29;23(6):e24435 [FREE Full text] [doi: 10.2196/24435] [Medline: 34115608]
27. Statista. URL: https://de.statista.com/ [accessed 2021-03-04]
28. Vidal-Alaball J, Fernandez-Luque L, Marin-Gomez FX, Ahmed W. A new tool for public health opinion to give insight into telemedicine: Twitter poll analysis. JMIR Form Res 2019 May 28;3(2):e13870 [FREE Full text] [doi: 10.2196/13870] [Medline: 31140442]
29. Eibensteiner F, Ritschl V, Nawaz FA, Fazel SS, Tsagkaris C, Kulnik ST, et al. People's willingness to vaccinate against COVID-19 despite their safety concerns: Twitter poll analysis. J Med Internet Res 2021 Apr 29;23(4):e28973 [FREE Full text] [doi: 10.2196/28973] [Medline: 33872185]
30. Yousefinaghani S, Dara R, Mubareka S, Papadopoulos A, Sharif S. An analysis of COVID-19 vaccine sentiments and opinions on Twitter. Int J Infect Dis 2021 Jul;108:256-262 [FREE Full text] [doi: 10.1016/j.ijid.2021.05.059] [Medline: 34052407]
31. Germani F, Biller-Andorno N. The anti-vaccination infodemic on social media: a behavioral analysis. PLoS One 2021;16(3):e0247642 [FREE Full text] [doi: 10.1371/journal.pone.0247642] [Medline: 33657152]
32. Truong J, Bakshi S, Wasim A, Ahmad M, Majid U. What factors promote vaccine hesitancy or acceptance during pandemics? A systematic review and thematic analysis. Health Promot Int 2022 Feb 17;37(1):daab105. [doi: 10.1093/heapro/daab105] [Medline: 34244738]
33. Rimmer A. Covid vaccination to be mandatory for NHS staff in England from spring 2022. BMJ 2021 Nov 10;375:n2733. [doi: 10.1136/bmj.n2733] [Medline: 34758984]
34. Horowitz J, Eddy M. Austria announces covid vaccine mandate, crossing a threshold for Europe. The New York Times. URL: https://www.nytimes.com/2021/11/19/world/europe/austria-covid-vaccine-mandate-lockdown.html [accessed 2021-11-22]
35. Schuetze CF. Germany shuts unvaccinated people out of much of public life. The New York Times. URL: https://www.nytimes.com/2021/12/02/world/europe/germany-unvaccinated-restrictions.html [accessed 2021-12-04]
36. Semlitsch T, Blank W, Kopp I, Siering U, Siebenhofer A. Evaluating guidelines: a review of key quality criteria. Dtsch Arztebl Int 2015 Jul 06;112(27-28):471-478. [doi: 10.3238/arztebl.2015.0471] [Medline: 26214233]

37. Graeber D, Schmidt-Petri C, Schröder C. Attitudes on voluntary and mandatory vaccination against COVID-19: evidence from Germany. PLoS One 2021;16(5):e0248372 [FREE Full text] [doi: 10.1371/journal.pone.0248372] [Medline: 33970933]
38. Spitale G, Merten S, Jafflin K, Schwind B, Kaiser-Grolimund A, Biller-Andorno N. A novel risk and crisis communication platform to bridge the gap between policy makers and the Public in the Context of the COVID-19 Crisis (PubliCo): protocol for a mixed methods study. JMIR Res Protoc 2021 Nov 01;10(11):e33653 [FREE Full text] [doi: 10.2196/33653] [Medline: 34612823]
39. Digital Health and Patient Safety Platform. URL: https://digitalpatientsafety.com/ [accessed 2021-03-19]
40. Followerwonk. URL: https://followerwonk.com/ [accessed 2021-02-20]
41. Digital Health and Patient Safety Platform. Facebook. URL: https://www.facebook.com/DHPSP [accessed 2021-03-19]
42. Ludwig Boltzmann Institute for Digital Health and Patient Safety. LinkedIn. URL: https://at.linkedin.com/company/ludwig-boltzmann-institute-for-digital-health-and-patient-safety [accessed 2021-03-19]
43. Symplur. URL: https://www.symplur.com/ [accessed 2021-02-20]
44. Crutzen R, Ygram Peters G, Mondschein C. Why and how we should care about the General Data Protection Regulation. Psychol Health 2019 Nov;34(11):1347-1357. [doi: 10.1080/08870446.2019.1606222] [Medline: 31111730]
45. Mathieu E, Ritchie H, Ortiz-Ospina E, Roser M, Hasell J, Appel C, et al. A global database of COVID-19 vaccinations. Nat Hum Behav 2021 Jul;5(7):947-953. [doi: 10.1038/s41562-021-01122-8] [Medline: 33972767]
46. Schwarzinger M, Watson V, Arwidson P, Alla F, Luchini S. COVID-19 vaccine hesitancy in a representative working-age population in France: a survey experiment based on vaccine characteristics. Lancet Public Health 2021 Apr;6(4):e210-e221 [FREE Full text] [doi: 10.1016/S2468-2667(21)00012-8] [Medline: 33556325]
47. Rhodes A, Hoq M, Measey M, Danchin M. Intention to vaccinate against COVID-19 in Australia. Lancet Infect Dis 2021 May;21(5):e110 [FREE Full text] [doi: 10.1016/S1473-3099(20)30724-6] [Medline: 32941786]
48. Lin C, Tu P, Beitsch LM. Confidence and receptivity for COVID-19 vaccines: a rapid systematic review. Vaccines (Basel) 2020 Dec 30;9(1):16 [FREE Full text] [doi: 10.3390/vaccines9010016] [Medline: 33396832]
49. Hacquin AS, Altay S, de Araujo E, Chevallier C, Mercier H. Sharp rise in vaccine hesitancy in a large and representative sample of the French population: reasons for vaccine hesitancy. PsyArXiv Preprint posted online on November 16, 2020. [FREE Full text] [doi: 10.31234/osf.io/r8h6z]
50. Verger P, Scronias D, Dauby N, Adedzi K, Gobert C, Bergeat M, et al. Attitudes of healthcare workers towards COVID-19 vaccination: a survey in France and French-speaking parts of Belgium and Canada, 2020. Euro Surveill 2021 Jan;26(3):2002047 [FREE Full text] [doi: 10.2807/1560-7917.ES.2021.26.3.2002047] [Medline: 33478623]
51. Kelekar AK, Lucia VC, Afonso NM, Mascarenhas AK. COVID-19 vaccine acceptance and hesitancy among dental and medical students. J Am Dent Assoc 2021 Aug;152(8):596-603 [FREE Full text] [doi: 10.1016/j.adaj.2021.03.006] [Medline: 34030867]
52. Barello S, Nania T, Dellafiore F, Graffigna G, Caruso R. 'Vaccine hesitancy' among university students in Italy during the COVID-19 pandemic. Eur J Epidemiol 2020 Aug;35(8):781-783 [FREE Full text] [doi: 10.1007/s10654-020-00670-z] [Medline: 32761440]
53. Murphy J, Vallières F, Bentall RP, Shevlin M, McBride O, Hartman TK, et al. Psychological characteristics associated with COVID-19 vaccine hesitancy and resistance in Ireland and the United Kingdom. Nat Commun 2021 Jan 04;12(1):29. [doi: 10.1038/s41467-020-20226-9] [Medline: 33397962]
54. Tokish H, Solanto M. The problem of vaccination refusal: a review with guidance for pediatricians. Curr Opin Pediatr 2020 Oct;32(5):683-693. [doi: 10.1097/MOP.0000000000000937] [Medline: 32773578]
55. Hayes L, Pollock AM. Mandatory covid-19 vaccination for care home workers. BMJ 2021 Jul 07;374:n1684. [doi: 10.1136/bmj.n1684] [Medline: 34233895]
56. Franco JVA. Should COVID-19 vaccination be mandatory? BMJ Evid Based Med 2021 Dec;26(6):269-270. [doi: 10.1136/bmjebm-2021-111813] [Medline: 34711633]
57. Hennock EP. Vaccination policy against smallpox, 1835-1914: a comparison of England with Prussia and Imperial Germany. Soc Hist Med 1998 Apr;11(1):49-71. [doi: 10.1093/shm/11.1.49] [Medline: 11620154]
58. Vetrugno G, Cicconi M, Foti F, Spagnolo A, De-Giorgio F. Compulsory vaccination would exacerbate resistance. BMJ 2019 Jul 04;366:l4464. [doi: 10.1136/bmj.l4464] [Medline: 31272948]
59. Salmon DA, Teret SP, MacIntyre CR, Salisbury D, Burgess MA, Halsey NA. Compulsory vaccination and conscientious or philosophical exemptions: past, present, and future. Lancet 2006 Feb 04;367(9508):436-442. [doi: 10.1016/S0140-6736(06)68144-0] [Medline: 16458770]
60. Arie S. Compulsory vaccination and growing measles threat. BMJ 2017 Jul 20;358:j3429. [doi: 10.1136/bmj.j3429] [Medline: 28729469]
61. Gagneux-Brunon A, Botelho-Nevers E, Bonneton M, Peretti-Watel P, Verger P, Launay O, et al. Public opinion on a mandatory COVID-19 vaccination policy in France: a cross-sectional survey. Clin Microbiol Infect 2022 Mar;28(3):433-439 [FREE Full text] [doi: 10.1016/j.cmi.2021.10.016] [Medline: 34774756]
62. Gallè F, Sabella EA, Roma P, Da Molin G, Diella G, Montagna MT, et al. Acceptance of COVID-19 vaccination in the elderly: a cross-sectional study in southern Italy. Vaccines (Basel) 2021 Oct 21;9(11):1222 [FREE Full text] [doi: 10.3390/vaccines9111222] [Medline: 34835152]

63. Giannouchos TV, Steletou E, Saridi M, Souliotis K. Mandatory vaccination support and intentions to get vaccinated for COVID-19: results from a nationally representative general population survey in October 2020 in Greece. J Eval Clin Pract 2021 Aug;27(4):996-1003 [FREE Full text] [doi: 10.1111/jep.13588] [Medline: 34060686]
64. Largent EA, Persad G, Sangenito S, Glickman A, Boyle C, Emanuel EJ. US public attitudes toward COVID-19 vaccine mandates. JAMA Netw Open 2020 Dec 01;3(12):e2033324 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.33324] [Medline: 33337490]
65. Paul KT, Eberl J, Partheymüller J. Policy-relevant attitudes toward COVID-19 vaccination: associations with demography, health risk, and social and political factors. Front Public Health 2021;9:671896. [doi: 10.3389/fpubh.2021.671896] [Medline: 34295869]
66. Tahir MJ, Saqlain M, Tariq W, Waheed S, Tan SHS, Nasir SI, et al. Population preferences and attitudes towards COVID-19 vaccination: a cross-sectional study from Pakistan. BMC Public Health 2021 Sep 26;21(1):1759 [FREE Full text] [doi: 10.1186/s12889-021-11814-5] [Medline: 34565351]
67. Chou WS, Oh A, Klein WMP. Addressing health-related misinformation on social media. JAMA 2018 Dec 18;320(23):2417-2418. [doi: 10.1001/jama.2018.16865] [Medline: 30428002]
68. Schmidt AL, Zollo F, Scala A, Betsch C, Quattrociocchi W. Polarization of the vaccination debate on Facebook. Vaccine 2018 Jun 14;36(25):3606-3612. [doi: 10.1016/j.vaccine.2018.05.040] [Medline: 29773322]
69. Dunn AG, Leask J, Zhou X, Mandl KD, Coiera E. Associations between exposure to and expression of negative opinions about human papillomavirus vaccines on social media: an observational study. J Med Internet Res 2015 Jun 10;17(6):e144 [FREE Full text] [doi: 10.2196/jmir.4343] [Medline: 26063290]

Abbreviations

DHPSP: Digital Health and Patient Safety Platform

Edited by A Mavragani; submitted 16.12.21; peer-reviewed by M Alfelali, G Spitale, G Kernohan, C Weerth; comments to author 08.01.22; revised version received 27.01.22; accepted 19.04.22; published 21.06.22.

Please cite as:
Ritschl V, Eibensteiner F, Mosor E, Omara M, Sperl L, Nawaz FA, Siva Sai C, Cenanovic M, Devkota HP, Hribersek M, De R, Klager E, Schaden E, Kletecka-Pulker M, Völkl-Kernstock S, Willschke H, Aufricht C, Atanasov AG, Stamm T
Mandatory Vaccination Against COVID-19: Twitter Poll Analysis on Public Health Opinion
JMIR Form Res 2022;6(6):e35754
URL: https://formative.jmir.org/2022/6/e35754
doi:10.2196/35754
PMID:35617671

Original Paper

System for Context-Specific Visualization of Clinical Practice Guidelines (GuLiNav): Concept and Software Implementation

Jonas Fortmann[1,2], MSc; Marlene Lutz[1], MSc; Cord Spreckelsen[2,3], Prof Dr

[1]Institute of Medical Informatics, Medical Faculty, Rheinisch-Westfälische Technische Hochschule Aachen University, Aachen, Germany

[2]Smart Medical Technology for Healthcare Consortium of the German Medical Informatics Initiative, Leipzig, Germany

[3]Institute of Medical Statistics, Computer and Data Sciences, Jena University Hospital, Jena, Germany

Corresponding Author:
Jonas Fortmann, MSc
Institute of Medical Informatics
Medical Faculty
Rheinisch-Westfälische Technische Hochschule Aachen University
Pauwelsstrasse 30
Aachen, 52074
Germany
Phone: 49 024 1808 0260
Email: jonas.fortmann@rwth-aachen.de

Abstract

Background: Clinical decision support systems often adopt and operationalize existing clinical practice guidelines leading to higher guideline availability, increased guideline adherence, and data integration. Most of these systems use an internal state-based model of a clinical practice guideline to derive recommendations but do not provide the user with comprehensive insight into the model.

Objective: Here we present a novel approach based on dynamic guideline visualization that incorporates the individual patient's current treatment context.

Methods: We derived multiple requirements to be fulfilled by such an enhanced guideline visualization. Using business process and model notation as the representation format for computer-interpretable guidelines, a combination of graph-based representation and logical inferences is adopted for guideline processing. A context-specific guideline visualization is inferred using a business rules engine.

Results: We implemented and piloted an algorithmic approach for guideline interpretation and processing. As a result of this interpretation, a context-specific guideline is derived and visualized. Our implementation can be used as a software library but also provides a representational state transfer interface. Spring, Camunda, and Drools served as the main frameworks for implementation. A formative usability evaluation of a demonstrator tool that uses the visualization yielded high acceptance among clinicians.

Conclusions: The novel guideline processing and visualization concept proved to be technically feasible. The approach addresses known problems of guideline-based clinical decision support systems. Further research is necessary to evaluate the applicability of the approach in specific medical use cases.

(JMIR Form Res 2022;6(6):e28013) doi:10.2196/28013

KEYWORDS

clinical practice guideline; clinical decision support system; decision support techniques; computer-assisted decision making; guideline representation; workflow control patterns; workflow; clinical; decision making; support systems; software; eHealth; electronic health

Introduction

Clinical Practice Guidelines and Computerized Decision Support

Quality of care benefits from the application of clinical practice guidelines (CPGs) [1,2]. Various guideline-based clinical decision support systems (CDSSs) have been designed and implemented in the past [3]. Such CDSSs have been shown to improve adherence to CPGs, thus potentially increase quality of care [4].

Most guideline-based support systems are designed to modify the routine clinical workflow in a way that requires the clinician to interact with the CDSS directly at specific points in time to obtain additional information or other kinds of guidance with respect to the current treatment. These systems usually encompass a model of the underlying clinical pathway of some intervention and locally intervene at specific steps to provide additional information or recommendations.

Even though the respective research spans several decades and encompasses various CDSS implementations of different shapes, CDSSs are only recently becoming a common part of clinical practice. Several reviews addressed potential factors influencing the success or failure of guideline-based CDSS [5-8]. Among others, insufficient understanding of the underlying clinical processes turned out to be a relevant factor. Most CDSSs interact with the user by providing recommendations, reminders, or notifications to be considered at the specific point in time when they are shown. Greenes et al [5] propose that this focus of the interaction of the CDSS with individual decisions and actions should be complemented by a perspective that also considers the entire workflow. We also noticed the need for more attention to the overarching process when participating in the development of a guideline-based CDSS to assist in the treatment of bloodstream infections [9]. In this medical use case, the clinical workflow is executed over a relatively long period of time (up to 10 days). Thus, the impact of point-in-time decisions on the future workflow is of particular importance.

Internally, most systems contain computerized CPG knowledge, but they do not reveal it to the physician as a timewise longitudinal view of a CPGs intended treatment process. Only the current decision is shown and justified. Information about how this decision will influence the future clinical workflows is not presented (nor how past decisions have impacted the clinical workflow until now).

On one hand, the focus on the current decision to be made or the next steps to be taken is prioritized over the long-term perspective. On the other hand, visualizing the entire guideline (often hundreds of pages of condensed information) would understandably lead to an unacceptable cognitive load.

In this article, we present GuideLine Navigator (GuLiNav), an approach that addresses this shortcoming by generating context-specific versions of otherwise static clinical workflow representations. GuLiNav does so by avoiding a strict distinction between the design view of a computer-interpretable guideline (CIG; global representation of the guideline structure) and the view of the guideline during execution (presentation of the actual state/user interaction at a certain point in the patient journey), which is prevalent in most previous approaches.

The context-specific guideline generated by GuLiNav represents a feasible compromise between reducing the clinician's cognitive load (by reducing the information load presented to a manageable amount) and providing an overview of the global treatment situation (by preserving context-specific relevant information). It is meant to support the autonomy of the clinician by providing guideline-based advice that has been tailored to fit the context-specific circumstances.

State of the Art

Apart from CIG-related research, previous results in the area of presentation of treatment histories and timelines are also related to the system we present. For example, Plaisant et al [10] developed the LifeLines system, which intends to visualize personal histories of individuals and as such can also be used in the medical domain. It starts with an overview of the entire history and facilitates zooming in and out to reveal information on various levels of granularity. This kind of research lays a foundation for adequate digital representation of personal timelines (from a human-computer interaction perspective). More recently, several newer approaches to the visualization of time-oriented clinical data specifically have been developed. The general motivation is to enhance the presentation of raw clinical data by using a knowledge base to apply medical knowledge for deriving patient-specific advice or abstract medical concepts from clinical data (eg, Shahar et al [11], Martins et al [12], and Klimov et al [13] have developed different visualization approaches).

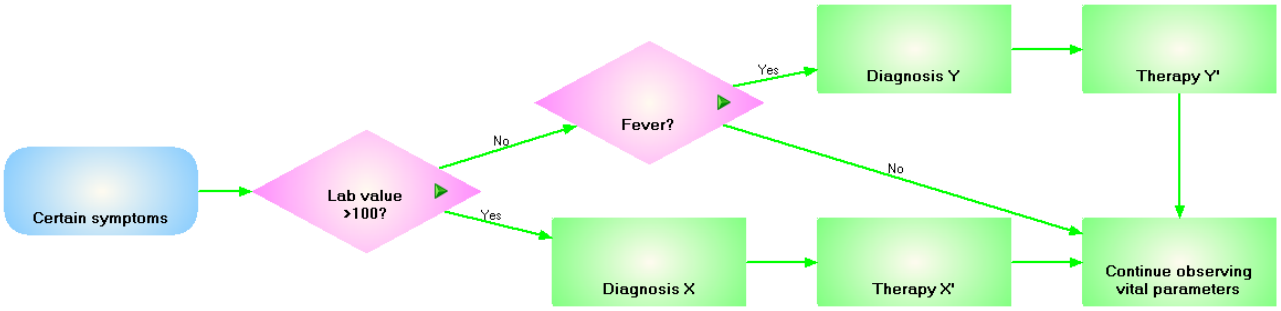
Within the last decades, various proprietary formats for the computerized representation of CPGs have evolved. They can be classified into 3 groups: document models, decision trees/probabilistic models, and task-network models (TNMs) [14]. The most prominent category with respect to its use in the context of clinical decision support are TNMs. Various TNMs were designed, such as the GuideLine Interchange Format, Version 3 (GLIF3), SAGE, GASTON framework, GLARE system, HELEN framework, PROforma formal knowledge representation language, Asbru, and more [15-21].

Even though the specific details vary, the basic idea is that in contrast to unstructured (ie, not machine readable) CPGs published by medical expert boards around the world, computer-interpretable formats facilitate the implementation of CDSSs by providing a structured, interchangeable, and, most importantly, computer-interpretable definition of a clinical guideline. This is achieved by representing the guideline in the form of some sequence of clearly defined tasks, actions, and decisions. During the treatment of a patient, patient-specific data can then be applied to the CIG to create an execution instance of the CIG and provide patient-specific advice.

For example, GLIF3 [15] uses 3 distinct levels to represent a CIG: conceptual, computable, and implementable. The conceptual level is represented as a flowchart (using a Unified Modeling Language [UML] Class Diagram) and acts as a structured documentation of the computerized guideline. Figure 1 shows the conceptual representation of an (imaginary) CIG.

The computable level has its own syntax and defines the data, actions, and algorithm flow, and at the implementable level, a GLIF-based CIG can be incorporated in a specific health information system.

Figure 1. An example computer-interpretable guideline modeled in GuideLine Interchange Format (GLIF) using the GLIF Editor from the Medical Objects Knowledgebase.



As another example, Asbru [21] uses a specialized syntax to specify so-called skeletal plans. Skeletal plans are, in short, temporal plans of procedures and actions captured on different levels of granularity. However, the Asbru syntax itself is defined in Backus-Naur form and thus difficult to read, even for domain experts. Thus, to visualize Asbru-based CIGs, AsbruView [22] was developed. It is a multidimensional representation that uses a traffic metaphor with different tracks representing different plans. The different axis then represents parallel plans or a decomposition of the underlying plan on a different level of detail.

In recent years, the notion that most of these specialized attempts are based on an equivalent foundation and can be, more or less, mapped from one to another, has emerged. This notion has surfaced outside of the medical domain: Russel et al [23] formally defined recurring patterns of workflows (using Petri Nets) and showed that the most prominent workflow systems, process models, and other related technical standard formats (such as SAP Workflow, FileNet, BPMN, UML 2.0 Activity Diagram, event-driven process chains) all express a large subset of the formally defined workflow control patterns and thus are, more or less, equivalent to each other.

Consequently, Mulyar et al [24] analyzed in which degree CIG representation can also be reduced to the same workflow control patterns. Despite the fact that a clinical guideline is inherently different from business workflows in the sense that it focuses on a single entity (the patient), they showed that CIGs also consist of basic workflow control patterns. They especially concluded that business process and model notation (BPMN), a prominent workflow modeling language from the business domain, has an equivalent expressiveness to the more specialized CIG formats.

More recently, guideline-based CDSSs have been developed using BPMN [25], often combined with ARDEN Syntax [26] modules, for the definition of the computerized guideline model. For example, de Bruin et al [27] used a combination of BPMN and ARDEN Syntax to model a clinical guideline for the prevention of transmission of hepatitis B from the mother to the newborn child. As another example, Rodriguez-Loya et al [28] used a combination of BPMN and a rule engine to diagnose chronic obstructive pulmonary disease as part of a workflow.

The computerized guideline representations are mostly intended as technical components to facilitate the implementation of guideline-based CDSSs but have no impact on what the actual end user of the CDSS sees.

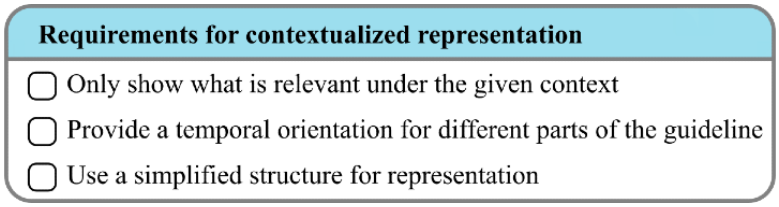
Past guideline representation approaches chose between 2 extremes: not taking the treatment context into consideration at all (eg, the guideline document itself or certain static pathways derived from it) or only presenting isolated single point-in-time decisions to the user without providing an overview of the workflow before and after that point in time. The first variant also presents information that is irrelevant with respect to a specific context (information overload) and consequentially also leads to a high cognitive load for the user. The second variant does not embed the current options into the larger scope of a treatment (what happened, what could happen in the future) and delegates these considerations to the user themself, who has to memorize this information. As a consequence, this kind of information underload also leads to a high cognitive load for the user. GuLiNav's contextualized guideline approach compromises between these 2 extremes.

Methods

Guideline Representation and Visualization

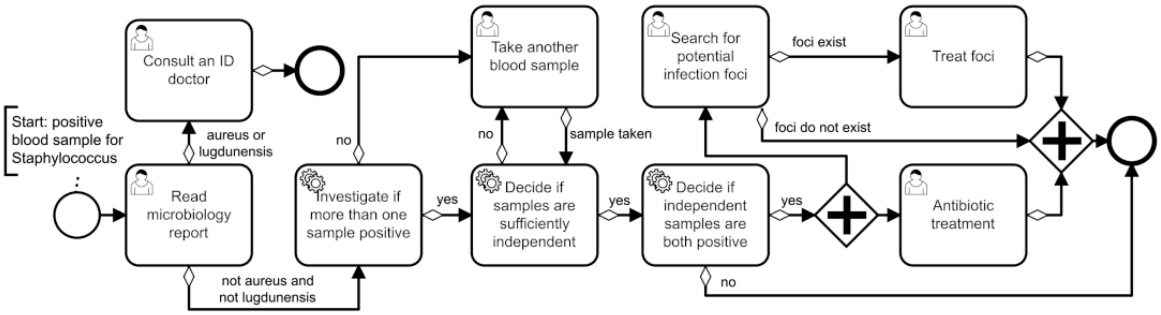
Clinical pathways are complex to define, but when being presented to humans, a schematic description usually suffices since humans are capable of filling in the information gaps using their implicit knowledge. However, the complexity increases significantly when the need for operationalization arises and every detail needs to be specified explicitly (compare, for example, with the qualitative study about lessons learned when implementing clinical decision support by Wright et al [29]). This makes it especially difficult to find a visual representation that is entirely machine readable and still somehow accessible to humans. We have identified 3 aspects for human-readable guideline visualization (see Figure 2) that we will explain in more detail in the further course of this article.

Figure 2. Short checklist of the requirements for contextualized representation.



We chose to use BPMN as a visual specification that represents the procedural aspects of the pathway (guideline procedures layer), while the medical criteria can be specified in an arbitrary fashion (eg, hard-coded, ARDEN Syntax...; medical criteria layer). This 2-way split of the guideline representation is a concept further explained and justified in Fortmann and Spreckelsen [30]. Other authors have suggested similar variants of separation between these two aspects (eg, Shahar et al [31] or Hatsek et al [32] suggest similar options on how to structure a computerized guideline). The decision to use BPMN is motivated by the fact that it can be used for the representation of CPGs [23,24,33], and there are also multiple useful tools, frameworks, and other resources freely available to work with it. Examples of these guidelines can be seen in Figure 3 and Figure 4. The examples are adaptations of the guideline used within the Hospital-Wide Electronic Medical Record Evaluated Computerized Decision Support System to Improve Outcomes of Patients With Staphylococcal Bloodstream Infection (HELP) study for the treatment of staphylococcal bloodstream infections [9], which initially motivated this research. Note that the example is simplified to be a fitting example for the rather technical focus of this article. The procedural view (visualized in BPMN) should be dynamized by taking an individual treatment context into consideration.

Figure 3. Context-based guideline visualization—Overview: Given a guideline definition and a treatment context (left), a context-sensitive guideline representation is generated (right).



Parameter	Value
Identified pathogen	Staph. epidermis
ID doctor consulted	?
More than one sample positive	?
Another blood sample taken	?
Samples sufficiently independent	?
Both independent samples positive	?
Infection foci	?
Foci treated	?
Antibiotic treatment	?

Read microbiology report	PAST
Consult an ID doctor	UNREACHABLE
Investigate if more than one sample positive	PRESENT
Decide if samples are sufficiently independent	FUTURE
Take another blood sample	FUTURE
Decide if independent samples are both positive	FUTURE
Search for potential infection foci	FUTURE
Treat foci	FUTURE
Antibiotic treatment	FUTURE

Ethics Approval

The project uses only synthetic, generated data for testing. At no point in time was any real patient data used. The evaluation was done with nonpatients (colleagues from the university hospital). In similar former cases, especially evaluation studies including participants from the university (students, staff), the ethical committee of RWTH Aachen University Hospital was consulted and showed no apprehension about the studies. It declared that its consent was not required (eg, Ethical IDs: 2019: 269/18, 2020: 270/18).

Context-Specific Guideline Visualization

Identification of 3 Aspects for Concept-Specific Visualization

Here we use the concept of a context-specific visualization. By context-specific we mean that clinical and patient data from a given situation are considered algorithmically to generate a guideline visualization adapted to the respective circumstances.

We identified 3 aspects where the context-specific visualization of a guideline should exploit context information to make the visualization more applicable (see Figure 2). Furthermore, we specified how each of these aspects can be considered when generating the context-sensitive guideline representation. As described in the introduction, the overarching goal is the

prevention of cognitive overload by shifting away from decision support for point-in-time decisions (to prevent information underload). Instead, we want decision support for perspective decisions that underlines the entire treatment history and points to possible directions for treatment while eliminating irrelevant information (to prevent information overload).

Show Only What Is Relevant Under the Given Context (Pruning)

Pathways usually contain forks that, depending on some external information, indicate in which direction the flow will continue. These kinds of if-then-else conditions usually cause entire parts of a pathway to become irrelevant once all information needed to make the decision is available. Obviously, this information can change and reevaluation is necessary whenever we see a change in relevant data. Thus, the unreachable part of the static definition should be conditionally pruned and not shown in the contextualized representation.

Provide a Temporal Orientation for Different Parts of the Guideline (Temporal Assignment)

During different points in time during the execution of a clinical pathway, specific tasks are performed in a certain order. Generally, there can be more than one formally correct order and more than one task can be active at the same time. In addition, the same task can be active multiple times (circles). Nonetheless, we can assign a temporal role to each task that describes if the task was already executed (and should not be executed again; PAST) or is a choice that can currently be chosen (PRESENT) or might possibly be executed at a later time (FUTURE). Thus, the temporal role of each task should be visualized (ie, by assigning a color to each temporal role and then coloring each task respectively). Note that the temporal role PAST is only assigned if the task cannot occur again, and the temporal role FUTURE is also assigned to tasks that were already executed if they can potentially be executed again in the future.

Simplify the Structure for Representation (Topological Sorting)

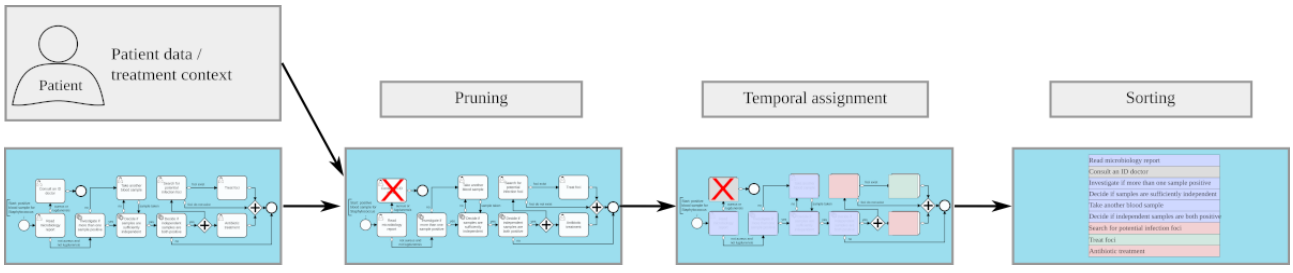
Since only relevant parts of the pathway are shown and other branches of the pathway are pruned, the context-sensitive pathway already has a simpler structure. However, the structure should be further simplified to reduce the mental workload necessary for a human to process the visualization. Since the pathway has already been pruned, a list representation would not removing as much information. Thus, the pruned pathway visualization should be further simplified by restructuring it into a list that contains the pathway's tasks in an intuitive, natural order.

For the third aspect, we needed a deterministic definition of what we consider a natural order. We decided to use the definition of a topological order from graph theory to find such an order. We used the default notation for graphs in theoretical computer science as, for example, used by Gibbons [34]. A topological order in graph theory is an ordering of the vertices of a directed acyclic graph where for every edge between 2 vertices u and v, the vertex u comes before the vertex v in the ordering.

Obviously, the graph often contains circles. To be able to apply a topological sorting algorithm, we internally remove the last edge from each circle. The last edge is the one that is the farthest from the starting vertex. The resulting order of the vertices, which represent procedures in our case, is one in which the procedures could potentially be executed.

Figure 3 shows a sketch that visualizes how such a context-sensitive visualization should look: given a guideline definition and context information (here: patient data), a pruned list of tasks (with assigned temporal role) can be created. Figure 4 illustrates the intermediate results after each of the 3 processing steps. To summarize, the context-specific guideline visualization provides support for the currently active tasks (point-in-time decisions) by identifying the currently relevant tasks (PRESENT). However, it also offers a simplified orientation over the entire treatment process by also showing past and (potential) future tasks and embedding the currently active tasks between them (perspective decisions).

Figure 4. Context-based guideline visualization—Processing steps: Intermediate results after each processing step during the generation of a context-sensitive guideline.

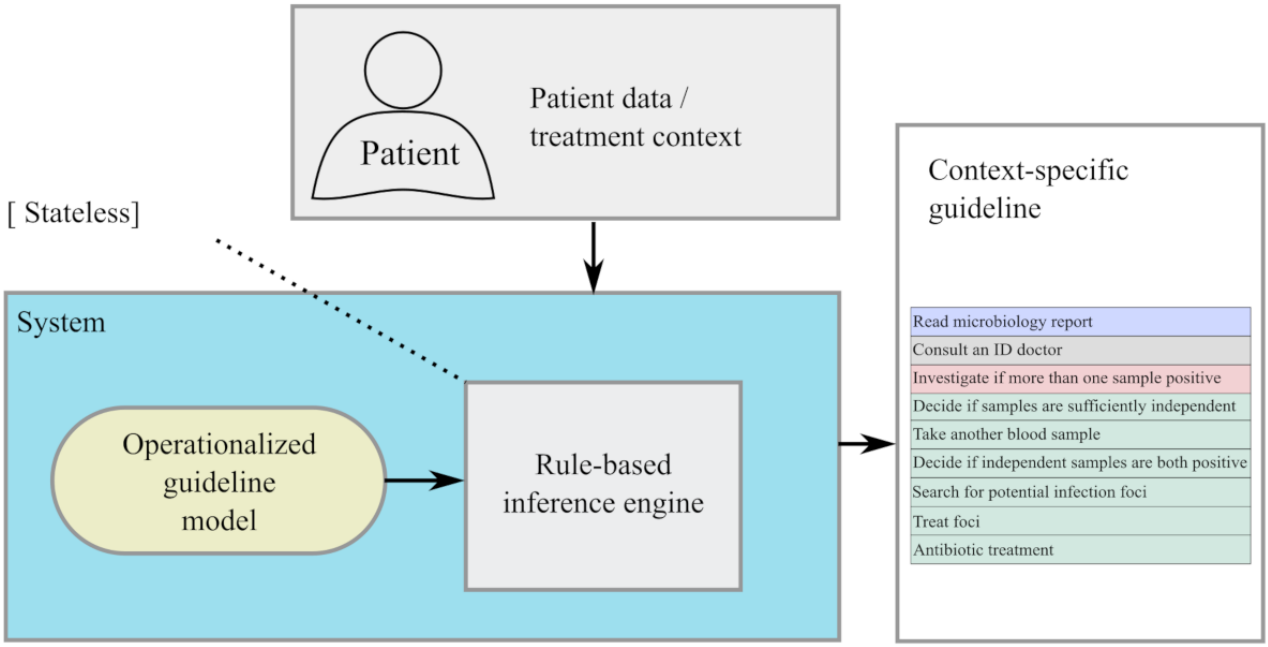


System Architecture

One problem with a treatment context is that a change in 1 parameter can turn the entire state of the clinical pathway upside down or even invalidate it (eg, in critical escalation scenarios). Thus, we designed the system in a stateless way. Each time the context changes (even in a minor way), the entire processing is redone. The system uses the knowledge specified in the operationalized guideline to infer a context-specific visualization each time a treatment context is given (see Figure 5). The prerequisite for this approach is that the full patient history is taken into account for each request. In this way, the contextualized guideline is constructed anew each time. Since the system produces a deterministic result, the context-specific

guideline will not change if the change in the context did not have any relevant impact on the guideline interpretation.

Figure 5. Schematic visualization illustrating stateless guideline processing by the system.



Internal CIG Representation

The CIG is processed by creating an enriched internal graph model. Start events, end events, gateways, and activities are the graph's vertices. A type property is assigned to each vertex describing which of the BPMN elements this vertex represents. For activities, the respective relevant properties of the treatment context are also stored. Sequence flows (ie, the transition elements of BPMN) are mapped to edges. For each edge, the original traversal condition from the sequence flow is stored.

CIG Interpretation

Use of a Business Rule Engine for Model Interpretation

The internal CIG model is interpreted using a business rule engine, a software system that executes business rules. These rules are typically structured in a "When PREDICATE then CONSEQUENCE" fashion. Business rule engines are closely related to the concept of logical programming. We created a set of business rules that define how the aforementioned internal CIG model shall be interpreted.

In the following sections, we will outline how these business rules are defined to infer a context-specific guideline representation that fulfills the 3 visualization aspects.

Pruning and Temporal Assignment

Given a CIG and a patient, the system first infers basic facts about the vertices and edges of the CIG. For example, for each vertex the rule system decides, by applying the patient data, if the activity was already performed or not. For each edge, it is decided if the condition is satisfied, unsatisfied, or cannot be decided with the data provided. These basic facts are then added into the rule system as predicates. The rules are then executed, inferring facts from the known predicates consecutively.

Eventually, a temporal role is inferred for each vertex, determined by the CIG's logic (gateway logic, conditional logic). Finally, the newly created facts are used to generate the contextualized guideline.

At this point, the contextualized guideline is still a graph, but vertices that are unreachable from the start vertex have already been pruned. While the shape of the context-specific guideline representation is conceptually determined by the 3 aspects already explained, the processing is only separated into 2 steps, since the pruning is implicitly performed during temporal assignment.

Topological Sorting

To obtain a linear structure from the graph that can be intuitively understood within the context used, a topological sorting algorithm is applied. The algorithm first removes circles in a way that preserves paths that begin in the starting vertex. Afterward, the algorithm is a modified version of Kahn's algorithm [35] that keeps vertices close to each other that are close in the original graph.

Results

Software Architecture

We have created a Java package that provides GuLiNav [36]. The app provides an internal Java application programming interface (API) and can be used as a software library within another Java project. Additionally, GuLiNav can run on its own and provides a representational state transfer (REST)-based interface that can be used to provide guideline models and context information to the system and in return provide context-specific guideline representations (see Figure 6).

Figure 6. Two distinct interfaces provided by the system: internal software library (Java API) or HTTP (representational state transfer API). CDSS: clinical decision support system; GuLiNav: GuideLine Navigator; API: application programming interface; REST: representational state transfer.

The inference of a context-specific representation is performed using a business rule engine. Note that the medical knowledge is not encoded using business rules. Rather, the business rules define how a CIG that is modeled using BPMN should be interpreted. An example can be seen in Figure 7. The criteria to be evaluated in each task can become arbitrarily complex and need to be evaluated in an additional layer that could use, for example, ARDEN syntax [26], but can be provided in arbitrary fashion (we, for example, defined a rather minimalistic software module where medical knowledge is encoded in an ARDEN medical logic modules–like fashion). The structure and purpose of the software facilitates software testing.

Figure 7. Exemplary Drools rules that are part of the rule system.

```
rule "ConditionSatisfied"
when
  $patient: GuidelinePatient()
  $edge: Edge(getCondition().evaluate($patient).isPresent() && getCondition().evaluate($patient).get())
then
  insert(new ConditionSatisfied($edge));
end

rule "PotentiallyConditionSatisfied"
when
  $patient: GuidelinePatient()
  $edge: Edge(getCondition().evaluate($patient).isPresent() == false)
then
  insert(new PotentiallyConditionSatisfied($edge));
end
```

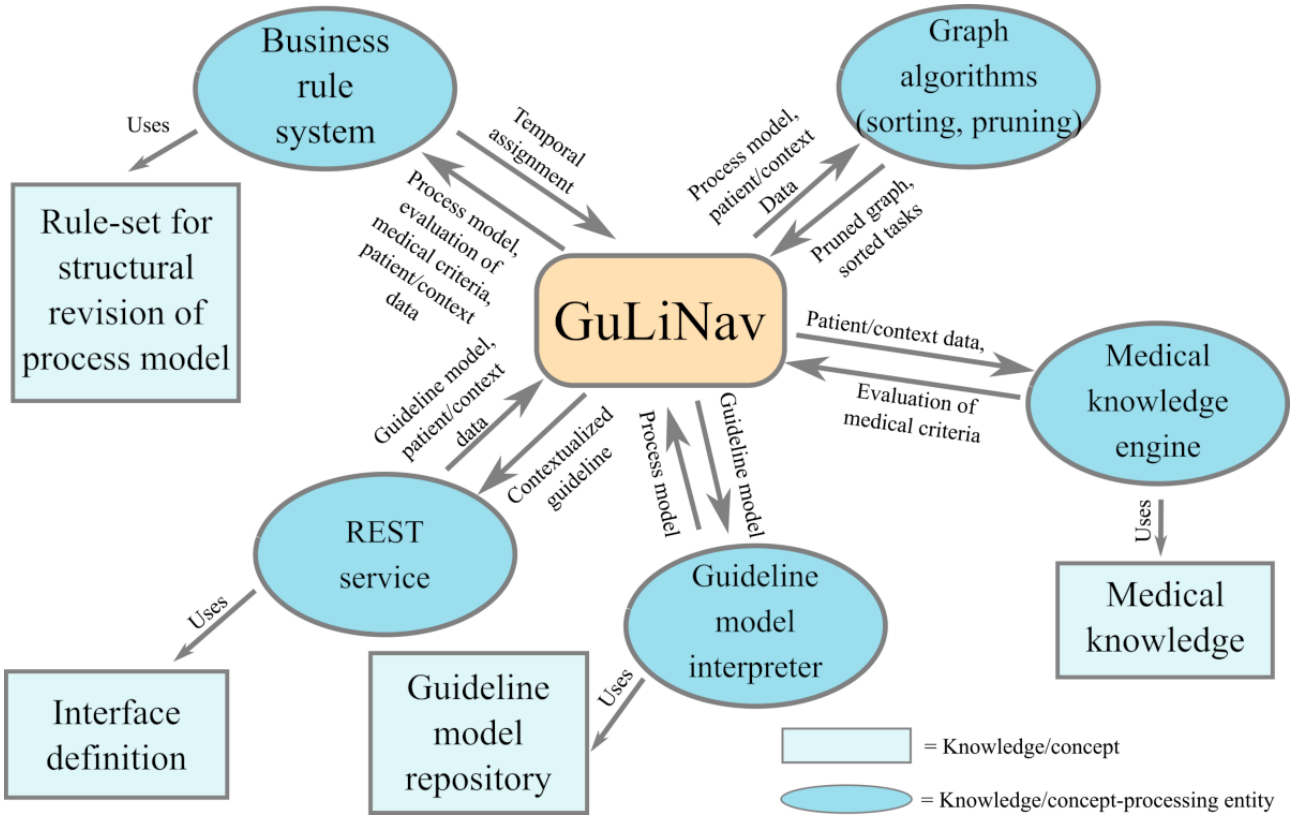
Frameworks Used

We used Spring Boot (2.1.8.RELEASE) [37] to create the REST-API. The rule-based inference is performed using the Drools (7.26.0.Final) [38] business rule engine. The BPMN model is read using the model API of the Camunda engine (7.11.0) [39]. Unified expressions are evaluated using the Java Unified Expression Language library (2.2.7) [40], and unit tests were defined using JUnit (4.4) [41]. For a proof of concept, we have also temporarily embedded the Arden2ByteCode compiler by Gietzelt et al [42] into GuLiNav and defined some exemplary ARDEN medical logic modules and used them for the evaluation of medical criteria.

Summary of the Concept and Software Implementation

To provide a summary for the concept and its technical implementation, Figure 8 shows an overarching diagram of different components and concepts used. The central component, GuLiNav, orchestrates the other modules of the system. As a communication point with the external software layers, a REST interface is provided (here: Spring Boot). External programs can use this interface to post new guideline models or request guideline contextualization of a previously posted guideline by posting patient and/or context data. The guideline model interpreter can provide process models of a previously posted guideline by interpreting the corresponding guideline model (here: Camunda BPMN). The medical knowledge engine is responsible for evaluation of the medical criteria layer as described in Fortmann and Spreckelsen [30]. That could be, for example, the aforementioned ARDEN engine, which uses encoded medical knowledge (eg, ARDEN MLMs). Pruning and topological sorting are performed by state-of-the-art graph algorithms directly implemented as part of GuLiNav. Finally, a business rule system (here: Drools) is used to infer the abstract temporal assignment of each of the guideline's tasks. It uses a rule set that defines how the temporal roles can be inferred from the combined information of the process model, evaluated medical criteria, and patient/context data. GuLiNav then eventually combines all the subsystem's responses to return a contextualized guideline via the REST interface.

Figure 8. Overarching diagram describing the relation between concepts and technologies used by the GuideLine Navigator. GuLiNav: GuideLine Navigator; REST: representational state transfer.



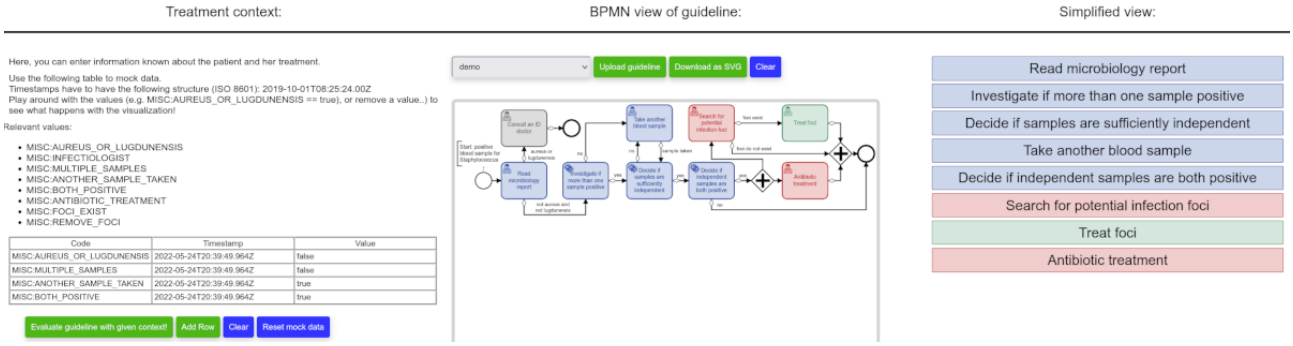
Knowledge Engineering

Knowledge acquisition for GuLiNav is performed by defining the procedural BPMN model first without completely specifying the medical criteria knowledge required in each task. Since BPMN is a widely adopted standard, there are many modeling tools available, and there is no need to define custom editors. We used Camunda BPMN [39] to model the procedural layer of the CIGs. BPMN is also rather easy to understand for nontechnicians. Thus, these procedural models can be discussed in interdisciplinary teams of information technology specialists and clinicians. The medical knowledge layer of the CIG can also be computerized by computer scientists by discussing individual knowledge modules in natural language with clinicians. These are then transcribed into the self-coded java package directly. The clear separation between procedural knowledge and medical criteria knowledge makes it possible to maintain these parts of the CIG separately, which causes the individual parts to remain relatively simple.

Demonstrator User Interface

We created a web-based front end for demonstration purposes. Figure 9 shows a screenshot of the demonstrator. Note that this graphical user interface additionally visualizes an intermediate result (pruned CIG before topological sorting) for debugging and demonstration purposes, which was the original reason for its implementation. This demonstrator can especially be used during interviews with domain experts: changes of the BPMN model can directly be posted to GuLiNav. The impact on the resulting context-specific guideline visualization is then immediately reflected, which enables a direct feedback loop between changes in the procedural model and the resulting context-specific guideline visualization, making it a useful tool for knowledge acquisition.

Figure 9. Screenshot of the system's demonstrator front end.



Technical Evaluation

On top of thorough technical testing of GuLiNav using unit tests, we have further validated the inference engine by exploiting the workflow control patterns of Russel et al [23]. Since the semantics of the guideline model are based on those patterns, we could validate the correct behavior of GuLiNav's inference engine by systematically defining test cases for all supported workflow control patterns. As an example, we present the test case for the synchronization pattern. It is described by Russel et al [23] as "the convergence of 2 or more branches into a single subsequent branch such that the thread of control is passed to the subsequent branch when all input branches have been enabled." The test case is shown in Table 1, and the pattern itself, modeled in BPMN, is shown in Figure 10. The system processes all possible combinations of patient data for value A and value B and the result is asserted accordingly. Task C does not always have a temporal role because in some cases it is pruned from the contextualized guideline.

Table 1. Test case definition for the synchronization pattern.

	Input		Expected inference		
Test-Pat-ID[a]	VAL[b] A	VAL B	TR[c] of A	TR of B	TR of C
Pat01	true	true	past	past	present
Pat02	true	—[d]	past	present	future
Pat03	true	false	past	past	—
Pat04	—	true	present	past	future
Pat05	—	—	present	present	future
Pat06	—	false	present	past	—
Pat07	false	true	past	past	—
Pat08	false	—	past	present	—
Pat09	false	false	past	past	—

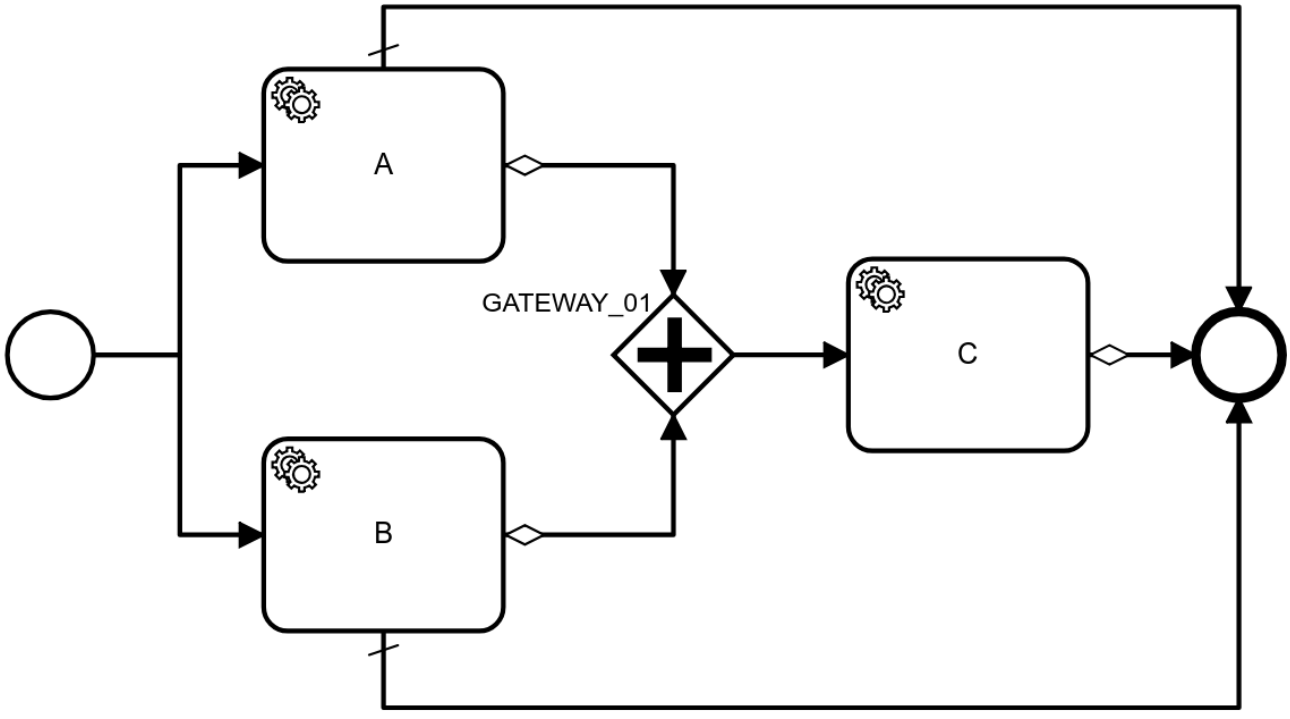
[a]Test-Pat-ID: test patient ID.

[b]VAL: value for task.

[c]TR: temporal role.

[d]Not present.

Figure 10. The business process model and notation model used for the test case of the synchronization pattern.



Formative Usability Evaluation

GuLiNav is a framework intended to be used by specific CDSS use cases and as such does not provide a user interface itself. It provides a linear data structure that can be used by the user interface to draw a context-specific guideline just by adding graphical elements. The concrete design in which this guideline representation is shown to the user can thus differ between use cases. We nonetheless executed a formative usability evaluation at this early stage by creating a mobile demonstrator based on the GuLiNav approach, which processes a guideline for the treatment of acute respiratory distress syndrome and presents a context-specific visualization of it to the user. The app was given to 6 clinicians on an iPad and the think-aloud protocol method was used for evaluation [43]. They could manipulate the patient's data and track the resulting changes in the context-specific guideline visualization. The general feedback toward the mobile app was mixed, and the think-aloud protocol revealed some usability issues. However, the concept of a context-specific guideline visualization in particular was positively received. It was intuitively understood and considered useful by the participants. The complete mobile app was subsequently evaluated using a questionnaire for software ergonomics, but for the context-specific guideline visualization in particular we have, until now, only collected the respective qualitative feedback.

Discussion

Principal Findings

GuLiNav shows the technical feasibility of combining CIGs with context information to infer context-specific guideline visualization which avoids cognitive overload while preserving an overview of the global treatment situation. It follows the 3-layer concept [30], in which the procedural aspects of the guideline are encoded using BPMN while the medical knowledge and criteria can be specified using an arbitrary system (eg, ARDEN syntax). It is possible to create such a system as a REST-based service that can then be consulted by other systems whenever needed.

Previous CIG approaches mostly focus on machine interpretability of the structured guideline model with the intention of it being used as a technical component of (multiple) guideline-based CDSSs [14]. They are designed to simplify the software implementation of guideline-based CDSSs and are not intended to be part of the respective system's user interface. The processing and visualization concept presented in this paper, in contrast, focuses on improving the visualization of the guideline itself. By applying context-specific data to a structured guideline format, it provides a context-specific (human readable) representation. It prevents overwhelming cognitive load and restrictive, inflexible workflows at the same time and thus addresses known issues of CIGs [5,6,8].

It should be noted that the simplifications made when contextualizing a guideline remove procedural knowledge from the original guideline. The result is a linear structure where, for example, no distinction is made between sequential and parallel tasks. Clinicians should generally be able to compensate for this information loss with their implicit knowledge (balance between knowledge in the world and knowledge in the head [44]), but further investigations are necessary to provide evidence. The linear guideline structure is well suited to be displayed on mobile devices with limited screen size. We think this is a major advantage for the development of mobile CDSSs.

Limitations

This visualization concept is well defined, and clinicians responded positively to it within a formative usability evaluation. However, more detailed evaluations with regard to how well clinicians understand the context-specific guideline depiction (with training or without training) and how much benefit it provides in clinical practice have yet to be performed.

From a technical viewpoint, this implementation demonstrates GuLiNav's method of operation, but only the most important workflow control patterns (compare with the workflow control patterns as defined by Russel et al [23]) are currently supported for the guideline procedures layer. The medical criteria layer has been implemented as a simple placeholder and is not yet considered in the resulting context-specific visualization.

Future Work

We intend to use GuLiNav as the back end of (potentially mobile) guideline-based CDSS apps. In the future, the system will be implemented as a generic web service and accessed via a standardized interface. We already began implementing a Fast Healthcare Interoperability Resources (FHIR) interface [45] to provide patient data to the system. The extent to which FHIR's clinical reasoning module provides the capabilities necessary to potentially also represent the respective static as well as contextualized guidelines in a standardized way has yet to be evaluated.

Even though the 2-way split [30] of the guideline is cleanly separated in the internal implementation of GuLiNav, currently only the procedural layer is presented in the context-specific guideline. This should be addressed in the future, since the guideline knowledge about medical concepts should also be accessible to the end user. One could imagine being able to tap into the items of a context-specific guideline to view the underlying medical concepts. This approach could match well with limited display sizes, which are predominant due to increased use of mobile devices.

Managing the complexity of medical algorithms and guidelines is a core challenge for the successful establishment of guideline-based CDSSs. In addition, existing CDSS-related research projects (such as the study by Hagel et al [9] to use a CDSS for the treatment of bloodstream infections) stressed the need—and critical effort—to obtain regulatory approval as a certified medical device (especially under the Medical Device Regulation). Establishing organizational as well as technical structures to qualify for the respective regulatory approval is inevitably necessary to use the concept in clinical practice. In the future, we plan to use context-specific guideline visualizations within a decision support to be used in clinical practice to evaluate the applicability of the approach in a practical setting.

Conclusions

Long-term effects and impact on the overarching clinical workflow should be given more attention when working with CDSSs [5]. We approached this proposition by developing GuLiNav, a system that prepares context-specific guideline visualizations aiming at reducing cognitive load while preserving orientation. GuLiNav, in its current form, demonstrates the technical feasibility.

The idea for a contextualized guideline visualization emerged during the early stages of the development of a guideline-based CDSS for the treatment of specific bloodstream infections [9]. The context-specific visualization concept was evaluated as part of a formative usability test, and clinicians generally approved of it. However, further research in actual clinical settings is necessary to better estimate the applicability and usefulness of the approach.

Acknowledgments

The project reported here was partially supported by grant 01ZZ1803B from the German Federal Ministry of Education and Research in the context of the Smart Medical Information Technology for Healthcare consortium.

Authors' Contributions

Jonas Fortmann implemented most of the software and wrote the manuscript. Marlene Lutz implemented the demonstrator front-end and provided feedback on the manuscript, Cord Spreckelsen provided the initial idea, gave feedback on the software during all stages of development and revised the manuscript.

Conflicts of Interest

None declared.

References

1. Lugtenberg M, Burgers JS, Westert GP. Effects of evidence-based clinical practice guidelines on quality of care: a systematic review. Qual Saf Health Care 2009 Oct;18(5):385-392. [doi: 10.1136/qshc.2008.028043] [Medline: 19812102]
2. Grimshaw J, Russell I. Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. Lancet 1993 Nov 27;342(8883):1317-1322. [doi: 10.1016/0140-6736(93)92244-n] [Medline: 7901634]
3. Wright A, Sittig DF. A four-phase model of the evolution of clinical decision support architectures. Int J Med Inform 2008 Oct;77(10):641-649 [FREE Full text] [doi: 10.1016/j.ijmedinf.2008.01.004] [Medline: 18353713]
4. Murphy EV. Clinical decision support: effectiveness in improving quality processes and clinical outcomes and factors that may influence success. Yale J Biol Med 2014 Jun;87(2):187-197 [FREE Full text] [Medline: 24910564]
5. Greenes RA, Bates DW, Kawamoto K, Middleton B, Osheroff J, Shahar Y. Clinical decision support models and frameworks: seeking to address research issues underlying implementation successes and failures. J Biomed Inform 2018 Feb;78:134-143 [FREE Full text] [doi: 10.1016/j.jbi.2017.12.005] [Medline: 29246790]
6. Roshanov PS, Fernandes N, Wilczynski JM, Hemens BJ, You JJ, Handler SM, et al. Features of effective computerised clinical decision support systems: meta-regression of 162 randomised trials. BMJ 2013 Feb 14;346:f657 [FREE Full text] [doi: 10.1136/bmj.f657] [Medline: 23412440]
7. Sittig DF, Krall MA, Dykstra RH, Russell A, Chin HL. A survey of factors affecting clinician acceptance of clinical decision support. BMC Med Inform Decis Mak 2006 Feb 01;6:6 [FREE Full text] [doi: 10.1186/1472-6947-6-6] [Medline: 16451720]
8. Bates DW, Kuperman GJ, Wang S, Gandhi T, Kittler A, Volk L, et al. Ten commandments for effective clinical decision support: making the practice of evidence-based medicine a reality. J Am Med Inform Assoc 2003;10(6):523-530 [FREE Full text] [doi: 10.1197/jamia.M1370] [Medline: 12925543]
9. Hagel S, Gantner J, Spreckelsen C, Fischer C, Ammon D, Saleh K, et al. Hospital-wide electronic medical record evaluated computerised decision support system to improve outcomes of patients with staphylococcal bloodstream infection (HELP): study protocol for a multicentre stepped-wedge cluster randomised trial. BMJ Open 2020 Feb 10;10(2):e033391 [FREE Full text] [doi: 10.1136/bmjopen-2019-033391] [Medline: 32047014]
10. Plaisant C, Milash B, Rose A, Widoff S, Shneiderman B. LifeLines: visualizing personal histories. Proc SIGCHI Conf Hum Factors in Comput Syst 1996:221-227. [doi: 10.1145/238386.238493]
11. Shahar Y, Goren-Bar D, Boaz D, Tahan G. Distributed, intelligent, interactive visualization and exploration of time-oriented clinical data and their abstractions. Artif Intell Med 2006 Oct;38(2):115-135. [doi: 10.1016/j.artmed.2005.03.001] [Medline: 16343873]
12. Martins SB, Shahar Y, Goren-Bar D, Galperin M, Kaizer H, Basso LV, et al. Evaluation of an architecture for intelligent query and exploration of time-oriented clinical data. Artif Intell Med 2008 May;43(1):17-34 [FREE Full text] [doi: 10.1016/j.artmed.2008.03.006] [Medline: 18442899]
13. Klimov D, Shahar Y, Taieb-Maimon M. Intelligent visualization and exploration of time-oriented data of multiple patients. Artif Intell Med 2010 May;49(1):11-31. [doi: 10.1016/j.artmed.2010.02.001] [Medline: 20303245]

14. Peleg M. Computer-interpretable clinical guidelines: a methodological review. J Biomed Inform 2013 Aug;46(4):744-763 [FREE Full text] [doi: 10.1016/j.jbi.2013.06.009] [Medline: 23806274]
15. Boxwala AA, Peleg M, Tu S, Ogunyemi O, Zeng QT, Wang D, et al. GLIF3: a representation format for sharable computer-interpretable clinical practice guidelines. J Biomed Inform 2004 Jun;37(3):147-161 [FREE Full text] [doi: 10.1016/j.jbi.2004.04.002] [Medline: 15196480]
16. De Clercq PA, Blom JA, Hasman A, Korsten HH. GASTON: an architecture for the acquisition and execution of clinical guideline-application tasks. Med Inform Internet Med 2000;25(4):247-263. [doi: 10.1080/146392300455558] [Medline: 11198187]
17. Terenziani P, Montani S, Bottrighi A, Torchio M, Molino G, Correndo G. The GLARE approach to clinical guidelines: main features. Stud Health Technol Inform 2004;101:162-166. [Medline: 15537221]
18. Skonetzki S, Gausepohl H, van der Haak M, Knaebel S, Linderkamp O, Wetter T. HELEN, a modular framework for representing and implementing clinical practice guidelines. Methods Inf Med 2004;43(4):413-426. [Medline: 15472756]
19. Bury J, Fox J, Sutton D. The PROforma guideline specification language: Progress and prospects. Stud Health Technol Inform 2001;83:13-29. [doi: 10.3233/978-1-60750-927-1-13]
20. Tu SW, Campbell JR, Glasgow J, Nyman MA, McClure R, McClay J, et al. The SAGE Guideline Model: achievements and overview. J Am Med Inform Assoc 2007;14(5):589-598 [FREE Full text] [doi: 10.1197/jamia.M2399] [Medline: 17600098]
21. Miksch S, Shahar Y, Johnson P. Asbru: a task-specific, intention-based,time-oriented language for representing skeletal plans. Proc 7th Workshop Knowledge Eng Methods Lang 1997:9-19 [FREE Full text]
22. Miksch S, Kosara R, Shahar Y. AsbruView: visualization of time-oriented, skeletal plans. Proc Artif Intell Plann Syst Conf 1998:11-18 [FREE Full text]
23. Russell N, Ter Hofstede A, van der Aalst W, Mulyar N. Workflow control-flow patterns: a revised view. 2006. URL: http://workflowpatterns.com/documentation/documents/BPM-06-22.pdf [accessed 2022-05-15]
24. Mulyar N, van der Aalst WMP, Peleg M. A pattern-based analysis of clinical computer-interpretable guideline modeling languages. J Am Med Inform Assoc 2007;14(6):781-787 [FREE Full text] [doi: 10.1197/jamia.M2389] [Medline: 17712087]
25. Object Management Group: Business Process Model and Notation. URL: https://www.bpmn.org/ [accessed 2020-02-09]
26. Hripcsak G, Wigertz OB, Clayton PD. Origins of the Arden Syntax. Artif Intell Med 2018 Nov;92:7-9. [doi: 10.1016/j.artmed.2015.05.006] [Medline: 26254699]
27. de Bruin JS, Adlassnig K, Leitich H, Rappelsberger A. Separating business logic from medical knowledge in digital clinical workflows using business process model and notation and Arden syntax. Stud Health Technol Inform 2018;248:17-24. [Medline: 29726414]
28. Rodriguez-Loya S, Aziz A, Chatwin C. A service oriented approach for guidelines-based clinical decision support using BPMN. Stud Health Technol Inform 2014;205:43-47. [Medline: 25160142]
29. Wright A, Sittig DF, Ash JS, Erickson JL, Hickman TT, Paterno M, et al. Lessons learned from implementing service-oriented clinical decision support at four sites: a qualitative study. Int J Med Inform 2015 Nov;84(11):901-911. [doi: 10.1016/j.ijmedinf.2015.08.008] [Medline: 26343972]
30. Fortmann J, Spreckelsen C. Separating procedures and criteria in computerized clinical guidelines: a 3-layer approach. Stud Health Technol Inform 2019 Sep 03;267:118-125. [doi: 10.3233/SHTI190815] [Medline: 31483263]
31. Shahar Y, Young O, Shalom E, Galperin M, Mayaffit A, Moskovitch R, et al. A framework for a distributed, hybrid, multiple-ontology clinical-guideline library, and automated guideline-support tools. J Biomed Inform 2004 Oct;37(5):325-344 [FREE Full text] [doi: 10.1016/j.jbi.2004.07.001] [Medline: 15488747]
32. Hatsek A, Shahar Y, Taieb-Maimon M, Shalom E, Klimov D, Lunenfeld E. A scalable architecture for incremental specification and maintenance of procedural and declarative clinical decision-support knowledge. Open Med Inform J 2010 Dec 14;4(1):255-277 [FREE Full text] [doi: 10.2174/1874431101004010255] [Medline: 21611137]
33. Kaiser K, Marcos M. Leveraging workflow control patterns in the domain of clinical practice guidelines. BMC Med Inform Decis Mak 2016 Feb 10;16:20 [FREE Full text] [doi: 10.1186/s12911-016-0253-z] [Medline: 26863868]
34. Gibbons A. Algorithmic Graph Theory. Cambridge: Cambridge University Press; 1985.
35. Kahn AB. Topological sorting of large networks. Commun ACM 1962 Nov;5(11):558-562. [doi: 10.1145/368996.369025]
36. GuLiNav. URL: https://github.com/gulinav/_gulinav [accessed 2021-05-26]
37. Spring. VMware, Inc. URL: https://spring.io/ [accessed 2020-02-09]
38. Drools. Red Hat Inc. URL: https://www.drools.org/ [accessed 2020-02-09]
39. The Universal Process Orchestrator. Camunda. URL: https://camunda.com/ [accessed 2020-02-09]
40. Java Unified Expression Language. Odysseus Software GmbH. URL: http://juel.sourceforge.net/ [accessed 2020-02-09]
41. JUnit 5. URL: https://junit.org/ [accessed 2020-02-09]
42. Gietzelt M, Goltz U, Grunwald D, Lochau M, Marschollek M, Song B, et al. ARDEN2BYTECODE: a one-pass Arden Syntax compiler for service-oriented decision support systems based on the OSGi platform. Comput Methods Programs Biomed 2012 May;106(2):114-125 [FREE Full text] [doi: 10.1016/j.cmpb.2011.11.003] [Medline: 22206864]
43. Lewis C. Using the Thinking-Aloud Method in Cognitive Interface Design. Yorktown Heights: IBM TJ Watson Research Center; 1982.

44. Norman D. The Design of Everyday Things: Revised and Expanded Edition. New York: Basic Books; 2013.
45. Welcome to FHIR. URL: https://www.hl7.org/fhir/ [accessed 2021-05-14]

Abbreviations

API: application programming interface
BPMN: business process and model notation
CDSS: clinical decision support system
CIG: computer-interpretable guideline
CPG: clinical practice guideline
FHIR: Fast Healthcare Interoperability Resources
GLIF3: GuideLine Interchange Format, Version 3
GuLiNav: GuideLine Navigator
HELP: Hospital-Wide Electronic Medical Record Evaluated Computerized Decision Support System to Improve Outcomes of Patients With Staphylococcal Bloodstream Infection
REST: representational state transfer
TNM: task-network model
UML: Unified Modeling Language

Edited by A Mavragani; submitted 17.02.21; peer-reviewed by Y Shahar, B Middleton, G Myreteg, B Damoiseaux-Volman; comments to author 29.04.21; revised version received 14.06.21; accepted 17.03.22; published 22.06.22.

Please cite as:
Fortmann J, Lutz M, Spreckelsen C
System for Context-Specific Visualization of Clinical Practice Guidelines (GuLiNav): Concept and Software Implementation
JMIR Form Res 2022;6(6):e28013
URL: https://formative.jmir.org/2022/6/e28013
doi:10.2196/28013
PMID:35731571

Original Paper

Breast Cancer Physical Activity Mobile Intervention: Early Findings From a User Experience and Acceptability Mixed Methods Study

Gabriel Ruiz Signorelli[1], MSc; Francisco Monteiro-Guerra[2], PhD; Octavio Rivera-Romero[3], PhD; Francisco J Núñez-Benjumea[4], MSc; Luis Fernández-Luque[1], PhD

[1]Adhera Health, Inc, Palo Alto, CA, United States

[2]The Insight Centre for Data Analytics, School of Public Health, Physiotherapy and Sports Science, University College Dublin, Dublin, Ireland

[3]Department of Electronic Technology, Universidad de Sevilla, Seville, Spain

[4]Innovation Unit, Virgen Macarena University Hospital, Andalusian Health Service, Seville, Spain

Corresponding Author:
Luis Fernández-Luque, PhD
Adhera Health, Inc
Circuito II Bldg, Av de los Descubrimientos, s/n
Mairena del Aljarafe - Seville
Palo Alto, CA
United States
Phone: 1 656930901
Email: luis@adherahealth.com

Abstract

Background: Physical activity (PA) is the most well-established lifestyle factor associated with breast cancer (BC) survival. Even women with advanced BC may benefit from moderate PA. However, most BC symptoms and treatment side effects are barriers to PA. Mobile health coaching systems can implement functionalities and features based on behavioral change theories to promote healthier behaviors. However, to increase its acceptability among women with BC, it is essential that these digital persuasive systems are designed considering their contextual characteristics, needs, and preferences.

Objective: This study aimed to examine the potential acceptability and feasibility of a mobile-based intervention to promote PA in patients with BC; assess usability and other aspects of the user experience; and identify key considerations and aspects for future improvements, which may help increase and sustain acceptability and engagement.

Methods: A mixed methods case series evaluation of usability and acceptability was conducted in this study. The study comprised 3 sessions: initial, home, and final sessions. Two standardized scales were used: the Satisfaction with Life Scale and the International Physical Activity Questionnaire–Short Form. Participants were asked to use the app at home for approximately 2 weeks. App use and PA data were collected from the app and stored on a secure server during this period. In the final session, the participants filled in 2 app evaluation scales and took part in a short individual interview. They also completed the System Usability Scale and the user version of the Mobile App Rating Scale. Participants were provided with a waist pocket, wired in-ear headphones, and a smartphone. They also received printed instructions. A content analysis of the qualitative data collected in the interviews was conducted iteratively, ensuring that no critical information was overlooked.

Results: The International Physical Activity Questionnaire–Short Form found that all participants (n=4) were moderately active; however, half of them did not reach the recommended levels in the guidelines. System Usability Scale scores were all >70 out of 100 (72.5, 77.5, 95, and 80), whereas the overall user version of the Mobile App Rating Scale scores were 4, 4.3, 4.4, and 3.6 out of 5. The app was perceived to be nice, user-friendly, straightforward, and easy to understand. Recognition of achievements, the possibility of checking activity history, and the rescheduling option were positively highlighted. Technical difficulties with system data collection, particularly with the miscount of steps, could make users feel frustrated. The participants suggested improvements and indicated that the app has the potential to work well for survivors of BC.

Conclusions: Early results presented in this study point to the potential of this tool concept to provide a friendly and satisfying coaching experience to users, which may help improve PA adherence in survivors of BC.

(JMIR Form Res 2022;6(6):e32354) doi:10.2196/32354

KEYWORDS

breast cancer; BC; mobile app; physical activity; mHealth; acceptability; user experience; mobile phone

Introduction

Background

Breast cancer (BC) is the most prevalent diagnosed cancer in women worldwide [1] and the second leading cause of death in women [2,3]. Although BC affects a large and growing population of women worldwide [2,3], the survival rates are fortunately on a steady rise mainly because of advancements in screening and treatment [4]. BC is associated with a reduced quality of life (QoL) because of its symptoms and treatment side effects [5]. BC symptoms encompass both physical and psychological impairments. Physical symptoms and sequelae include loss of power and function in limbs, lymphedema, muscle wasting, loss of bone density, chronic fatigue, pain, weight gain, and loss of appetite, whereas psychological symptoms include depression, anxiety associated with uncertainty, poor body image, loss of intimacy in relationships, reduced self-esteem, and cognitive dysfunction [6,7]. Scientific evidence has demonstrated that physical activity (PA) is the most well-established lifestyle factor associated with BC survival [8]. PA provides vital benefits to patients with BC and survivors of BC, including prevention of cancer recurrence; decreased side effects from treatment; and improvements in fitness, body size, and QoL [8,9]. Even women with advanced BC may benefit from moderate PA [10]. However, most of the aforementioned symptoms and treatment side effects are barriers to PA, which may present actual or perceived risks of injury or discomfort during physical exertion [11]. In addition, women with BC frequently report other barriers to PA such as lack of time and information [12]. In such circumstances, participation in and adherence to PA is low among survivors of BC [13]. Some studies have reported estimates of <10% of survivors of BC, which meet the PA guidelines and recommendations [14]. Overcoming these barriers to PA adherence among women with BC, who meet the current recommendations, is a complex challenge that requires innovative and engaging strategies. There is growing evidence regarding coaching interventions that effectively engage women with BC in PA [15]. These interventions are based on behavioral change theories (BCTs) such as the social cognitive theory [16], transtheoretical model [17], and self-determination theory [18]. Often, the implementation of these techniques is negatively affected by a lack of engagement of the users with the technology used for the delivery of the interventions. In that regard, feasibility studies such as the one described in this paper can provide insights into potential barriers to digital behavioral interventions.

Information and communication technologies enable cost-effective alternatives to help people reach PA recommendations through digital BCT-based interventions. In particular, advancements in mobile health (mHealth) technologies have increased interest in the research and development of mobile PA coaching systems and interventions [19]. Digital health transformation is also increasing this interest, especially in the current global situation because of the COVID-19 pandemic in which patients are encouraged to take a proactive approach to the self-management of their health and QoL. Mobile devices present unique capabilities that enable data collection in real-life scenarios [20,21], just-in-time behavioral information provision [22], and remote communication with health care professionals. These capabilities allow remote assessment, tracking, and monitoring in real-time and real-life environments, which form the basis of momentary ecological interventions [23]. As a result, these mHealth systems may potentially empower patients, promote behavior changes toward a healthier lifestyle, facilitate self-monitoring of symptoms and behaviors [24], provide real-time tailored support and motivation [21], improve their educational level [25], and allow patients the feeling of being in contact with their health care team [26].

mHealth coaching systems take advantage of these capabilities to implement functionalities and features based on BCTs to promote healthier behaviors. Among these systems, digital PA coaching interventions are well-received by women with BC, as suggested in the increasing body of scientific evidence [12,27]. However, to increase its acceptability among women with BC, it is essential that these digital persuasive systems are designed considering their contextual characteristics, needs, and preferences [28]. Each woman experiences her cancer journey in a particular and dynamic way, receiving different treatments and experiencing diverse symptoms such as fatigue or pain, which affect her emotional state, reduce her physical and cognitive capacities, and demand personalized social support. This unique experience requires that digital PA coaching systems consider not only their needs at the group level but also tailored to each individual [11,27]. Previous studies aimed at investigating the specific requirements of women with BC for digital PA coaching interventions have highlighted the importance of personalization in adapting these interventions to the specific individual's conditions [12,27,29]. Lack of engagement and low perceived personal relevance of digital health systems are commonly associated with high levels of user abandonment [30]. In this sense, personalization can contribute to captivating and holding a person's interest [31], resulting in an increased long-term engagement and adherence to these systems. In addition, personalization is associated with an increase in the effectiveness of BCT-based systems [32]. There is a lack of knowledge regarding the technology acceptance of real-time behavioral feedback using mHealth technologies in survivors of BC.

Objective

This study presents the results of a small-scale evaluation of the acceptability of a personalized PA coaching mobile app for survivors of BC in the real world. The mobile app aimed to guide BC survivors on a plan to increase their PA, including behavioral and motivational aspects. The mobile solution also captured PA levels using a smartphone accelerometer. The objectives of this study were to (1) examine the potential acceptability or feasibility of the intervention; (2) assess usability and other aspects of the user experience; and (3) identify key considerations and aspects for future improvements,

which may help increase and sustain acceptability and engagement.

Methods

Study Design

A mixed methods case series evaluation of usability and acceptability was conducted in this study. The use of a mixed methods design was chosen to capture both quantitative feedback about the mobile solution engagement and qualitative feedback about the user experience. This approach is well suited for understanding areas of improvement before designing larger studies. This study draws from the theoretical framework of acceptability (TFA) developed by Sekhon et al [32]. The TFA states that acceptability is influenced by 7 dimensions: affective attitude, burden, ethicality, intervention coherence, opportunity costs, perceived effectiveness, and self-efficacy. There are examples of the use of TFA in health technologies [33,34]. The objective of this study was to understand each participant's perspective using both quantitative data (use of the solution) and qualitative data (interviews). A total of 4 women with BC participated in the study. The original plan was to recruit more participants; however, because of the COVID-19 pandemic, recruitment had to be halted.

The study comprised 3 sessions: initial, home, and final. In the initial session, a researcher provided participants with all the materials and instructions, installed the app on the smartphone, and explained the use of the system to the participants. In addition, data on participant characteristics were collected using questionnaires that covered demographics, technology use, and interests. In addition, 2 standardized scales were used: the Satisfaction with Life Scale [35] and the International Physical Activity Questionnaire–Short Form [36]. Satisfaction with Life Scale was used to capture health or life satisfaction as a potential mediator of overall engagement with the mobile solution.

Participants were then asked to use the app at home for a period of approximately 2 weeks in the home session, with data being collected from the app. App use and PA data were collected from the app and stored on a secure server during this period. The mobile solution was used to collect information on the PA and engagement of the users during this period. A period of 2 weeks was considered sufficient to capture the technology acceptance of the users. The 2 weeks were defined as sufficient time to identify major aspects related to acceptance of the solution, which was adapted to minimize disruptions in the clinical setting.

Finally, in the final session, participants returned materials to the research team that deleted any data stored on the smartphone and were asked to fill in 2 app evaluation scales and participate in a short individual interview. The participants completed the System Usability Scale [37] to assess the usability of the system. In addition, participants filled in the user version of the Mobile App Rating Scale [38] to assess the quality of the app. During the interviews, the interviewer provided trigger questions to the participants and took field notes. Questions used in the interviews were defined based on the TFA [32] and were built to cover various aspects of the participants' user experience [39] associated with the different dimensions of acceptability and usability. The interviews were audio recorded, transcribed, and anonymized. The initial and final sessions took place in a private room at the Beacon Hospital headquarters, where only the session facilitator and the participant were present, and lasted 35 to 60 minutes. The data were collected from February to the beginning of March 2020.

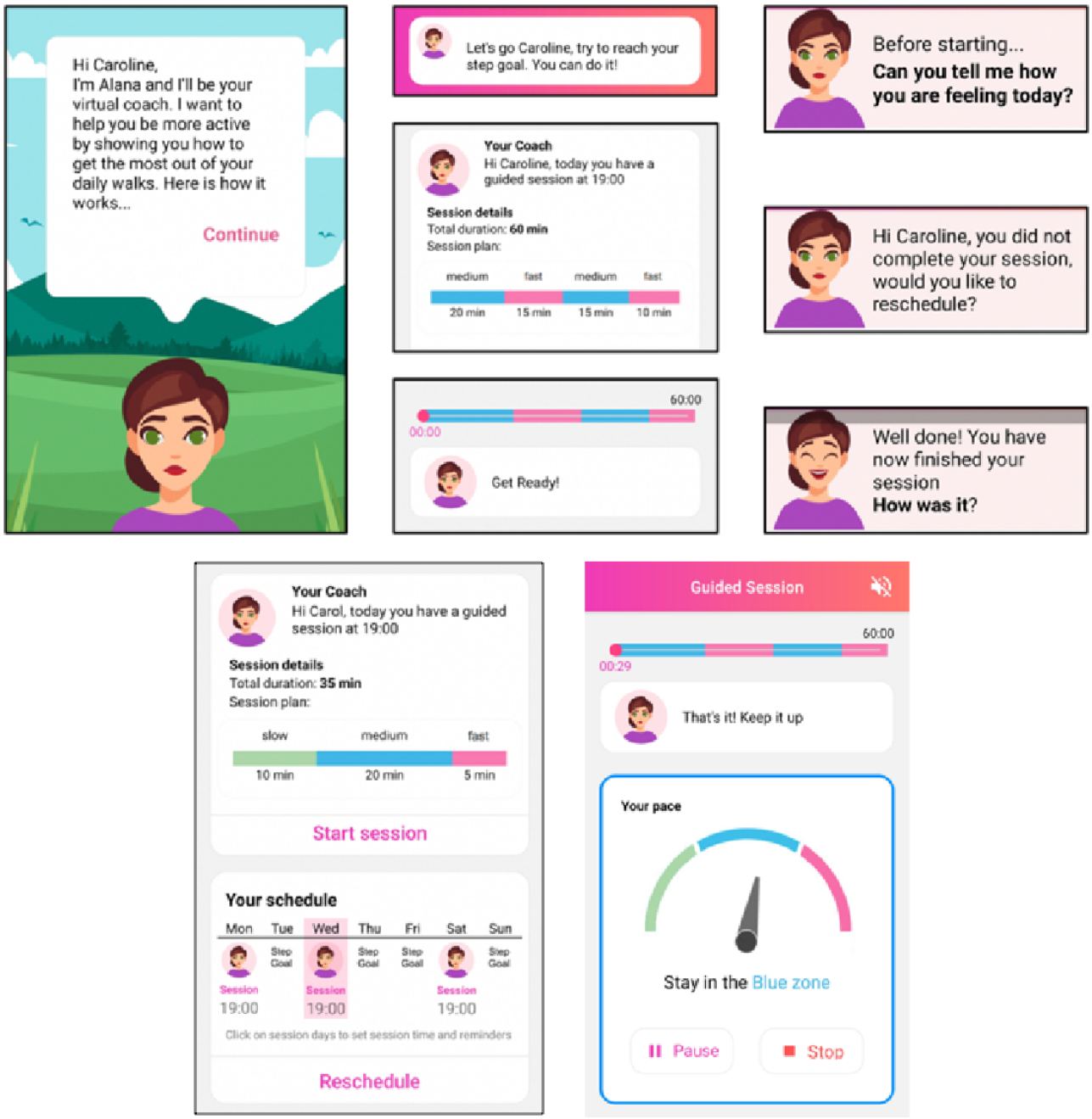
System and Materials

The proposed system aimed to function as a mobile personal PA coach for survivors of BC and focused on optimizing walking activities to help them reach and maintain the levels of PA recommended in the guidelines. A user-centered design approach was followed to ensure that the end user needs were met. Relevant theories and evidence for successful PA interventions were used as the basis for the system design process. A detailed description of the design process is published in the studies by Monteiro-Guerra et al [12,40]. Screenshots of the developed mobile app are shown in Figure 1. The main features considered for the developed system prototype were as follows:

- A walking regimen
- Real-time activity monitoring and feedback
- Real-time guided sessions (with instructions to control session intensity)
- Adaptative training
- Personalized and encouraging communication
- Interface simulating an app-based coach
- Activity scheduling tool and reminders
- Activity history
- Weekly summary reports

The participants were provided with a waist pocket (Kalenji, Decathlon), wired in-ear headphones (Ear Pollution Bolt, iFrogz), and an Android smartphone (P Smart, Huawei). The smartphone was provided to ensure a similar user experience across participants. They also received printed instructions. The waist pocket was flexible and adjustable in size. The participants were instructed to wear the waist pocket and carry the smartphone while completing their activities. In addition, participants were asked to wear headphones to optimize the audio feedback delivery, providing a more private experience, especially when activities were performed in public places.

Figure 1. Screenshots of the developed mobile app.



Recruitment

The study was conducted in collaboration with the Beacon Hospital (Ireland). Participants were recruited by the specialized oncology team based on information collected from the Beacon Hospital patient database and were eligible to participate if they (1) were patients of oncology with a history of BC who had finished primary curative treatment (surgery, radiotherapy, and chemotherapy), (2) owned and used a mobile phone or smartphone, (3) were able to speak and read English, (4) had no known impairments or comorbidities, and (5) had no restrictions on PA. The recruitment process was performed in 2 rounds. First, an email, including the participant information leaflet, was sent to potential participants; subsequently, eligible participants were contacted by a phone call. Participants were required to provide informed consent before participation.

Ethics Approval

This study was approved by the Research Ethics Committee of Beacon Hospital in Ireland (reference number: BEA0111) and received ethics exemption from the University College Dublin Human Research Ethics.

Data Analysis

Participants' PA levels were calculated based on their baseline session results, their mean number of steps per day, and the total number of sessions completed. Compliance with the training program was inferred considering three aspects: (1) the number of sessions completed of those proposed in the weekly plan, (2) the compliance with walking paces set in programmed sessions (equation 1), and (3) the number of nonsession days with the step goal achieved:

Here, R is the rate of compliance with the session format and N_s is the total number of sessions completed.

Data from the standardized scales were analyzed based on their standard procedures. The International Physical Activity Questionnaire–Short Form results were also used to estimate the adherence of participants to PA guidelines [23].

A content analysis of the qualitative data collected in the interviews was conducted iteratively, ensuring that no critical information was overlooked. The findings of the interviews were discussed among the authors across these iterations. Several categories were defined using the app evaluation scales and interview data to outline the findings: (1) perceived impact or benefit, (2) positive feelings about system features and aesthetics, (3) usefulness, (4) intervention feasibility and appropriateness, (5) usability, and (6) suggested improvements. On the basis of the TFA constructs and usability aspects, relevant categories and subcategories were compiled, highlighting both the individual and common perspectives across the 4 participants. NVivo (version 12; QSR International) software was used for content analysis. Key illustrative quotes were selected to highlight the perspective of each participant.

Results

Participant Characteristics

Participant characteristics, including contextual details that may have influenced their experience with the system and intervention, are presented in Table 1. The age of the participants ranged from 35 to 61 years. Of the 4 participants, 3 (75%) were highly educated, and all 4 (100%) were at least somewhat skilled and interested in technology. Time since diagnosis ranged from approximately 1 to 3 years, and 50% (2/4) of participants reported having some cancer-related physical limitations. Approximately 50% (2/4) of participants worked in an office, whereas the other 50% (2/4) were housewives. All participants were moderately active; however, 50% (2/4) did not reach the recommended levels in the guidelines. Of the 4 participants, 1 (25%) was single and reported being slightly dissatisfied with her life.

Table 1. Participant characteristics.

Characteristics	P1	P2	P3	P4
General characteristics				
Age (years)	61	43	35	54
Marital status	Married	Single	Married	Married
Education	Primary school	Postgraduate	Graduate	High school
Employment	Working (housewife)	Working (office)	Working (office+home)	Working (housewife)
Condition-related characteristics				
Date of diagnosis	February 2019	June 2018	October 2017	March 2018
Physical burdens	Fatigue and muscular pain	None reported	None reported	Lymphedema and joint pain
PA[a] level				
IPAQ-SF[b] score	Moderate	Moderate	Moderate	Moderate
Adherence to PA guidelines[c]	Not adherent	Adherent	Not adherent	Adherent
Sitting time (hours)	4	12	6	5
SWLS[d] result	Highly satisfied	Slightly dissatisfied	Highly satisfied	Satisfied
Technology use, interest, and skill				
Smartphone use	High	High	High	High
Interest in mobile technologies	Somewhat interested	Somewhat interested	Interested	Somewhat interested
Self-reported skill with technology	Somewhat skilled	Skilled	Skilled	Somewhat skilled
"I like to experiment with new technology"[e]	Somewhat agree	Somewhat disagree	Agree	Somewhat agree

[a]PA: physical activity.

[b]IPAQ-SF: International Physical Activity Questionnaire–Short Form.

[c]Adherence to >150 minutes per week=moderate activity or >75 minutes per week=vigorous activity, as inferred from the IPAQ-SF answers.

[d]SWLS: Satisfaction With Life Scale.

[e]Custom Likert scale made for the study.

Acceptability

The participants mentioned that the app was nice, user-friendly, straightforward, and easy to understand. They found the app useful and had a positive perception that it presented a training schedule to remind them that they had a target to achieve:

> *I liked the system. I think it's very well laid out.* [P1]
>
> *It's a lovely, easy app to use.* [P1]
>
> *It was straightforward to use.* [P2]
>
> *I thought it was a really nice, use-friendly app; it is very straightforward, I think it is easy to understand.* [P3]
>
> *I really like the app itself...* [P3]

Compliance

Training data on the initial level, intervention length, and compliance with the physical exercise program are shown in Table 2. One of the participants (P1) reported that she would have further followed the plan if it were not for bad weather and an ankle injury that she had experienced during the study period. Cold weather was also mentioned by P2. Regarding the training plan, some participants pointed out that some sessions were longer than the time they had available and that their current exercise habits were higher than what the app proposed:

> *On the first session of the week, [which is] 45 minutes, I would go, "Jesus, I only actually have time for 25 minute[s]," but you make the time for the 45 minutes and you feel good after it.* [P3]

Table 2. Intervention data overview.

Training aspect	P1	P2	P3	P4
Intervention length (weeks)	3	2	3	2.5
Total number of planned sessions, N	8	7	9	8
Total number of completed sessions, n (%)[a]	5 (63)	7 (100)	7 (78)	7 (88)
Daily step count, mean (SD)[b]	8251.7 (2742.9)	6838.3 (2335.9)	6377.4 (1942.3)	6292.4 (3575.8)
Training compliance				
With guided sessions (%)[a]	62.5	100	77.8	87.5
With step daily goal (%)[b]	66.7	28.6	70	60
Compliance with session format, mean (SD)[b]	97.7 (1.5)	98.6 (1.0)	95.0 (4.2)	87.8 (9.4)

[a]Includes sessions where the full data set was not recorded; however, participants reported having completed the session.

[b]Results calculated from sessions in which data were properly collected and stored for the duration of the session.

Usability

The results of the scores for the System Usability Scale and the user version of the Mobile Application Rating Scale are presented in Table 3. Participants mentioned that the technical difficulties with the system data collection, particularly with the miscount of steps, could make them feel frustrated. The system was somewhat cumbersome to use, given the need to use an extra phone for the study and the difficulty in carrying the phone while wearing the headphones:

> *If it was on my iPhone, I probably wouldn't think twice about using it.* [P4]
>
> *No, just bringing the (extra) phone with me. That's all.* [P4]
>
> *I put it in my pocket—it was grand. I always have zip pockets anyway. It was just in my pocket.* [P4]

Table 3. System Usability Scale (SUS) scores out of 100 and the mean user version of the Mobile App Rating Scale (uMARS) scores out of 5.

App evaluation scales or aspects	P1	P2	P3	P4
SUS	72.5	77.5	95	80
uMARS, mean (SD)				
Engagement	4 (0.7)	4 (0.7)	3.8 (1.3)	2.6 (1.1)
Functionality	3.8 (1.5)	4.5 (0.6)	4.75 (0.5)	4 (0.8)
Aesthetics	4.7 (0.5)	4.3 (0.6)	4.7 (0.6)	3.7 (0.6)
Information	3.8 (0.5)	4.5 (0.6)	4.5 (0.6)	4.3 (1.0)
Overall quality	4.0 (0.4)	4.3 (0.2)	4.4 (0.4)	3.6 (0.7)
App subjective quality	4.3 (1.0)	3.8 (0.5)	4.0 (0.8)	3.3 (1.7)
Perceived impact	4.7 (0.5)	3.0 (0.9)	4.0 (0.6)	3.5 (0.8)

About the functionalities of the app, the participants highlighted the recognition of achievements and the possibility to check their activity history and found that the rescheduling option was *excellent*. The system made them more aware and conscious of being active and stimulated them to be more active. Moreover, participants felt that they were not on their own. Guided sessions' cues to slow down and speed up and the progress in time through the session was important to encourage them to keep going, creating a feeling of satisfaction when they followed what the plan proposed:

> *I enjoyed it and I loved the fact that I could reschedule, because I could work it around the days and the times that suited me. The reschedule and the retiming I found excellent.* [P1]
>
> *I liked when she told you to speed up or slow down.* [P2]
>
> *I liked that she spoke to you and said, "We are now moving into this."* [P2]
>
> *You'd know—she'd say, "Five minutes more" or whatever—and that was good. I can imagine that's very encouraging. If I was running and she had been saying those things to me, I would have been encouraged to keep going. She'd say, "Only a few more minutes, or you're going faster than you need to." Yes.* [P2]

Rescheduling functionality was perceived to be very useful. Participants also had very positive perceptions toward being able to check their activity progress during the day and their past activity history, which they also found useful.

Finally, when questioned about battery consumption during the sessions, the participants stated that they did not perceive that it drained the battery.

Potential Improvements

Participants also suggested several improvements that they believed would make the system fit their preferences more effectively. These involved the inclusion of other health aspects, such as sleep and diet, in a diary to compare with the PA progress; combining the guided exercise with music, the possibility to pair the system with a wearable, receiving feedback on the sessions, and checking the number of steps taken; and that the pausing function should be easier, avoiding taking the phone out and pausing it at the traffic lights.

Participants also suggested that it would be nice to have more options to customize the training plan and sessions, with long-term goals (eg, to do a 5 km run), being able to customize the session distance or duration, or adding more sessions in a week and having the option to customize the voice of the coach. They also mentioned the idea of relating the pace zones with the heart rate to allow the user to customize the training plan and have a cooldown period in the guided sessions.

Regarding the history functionality, suggested improvements included having detailed information of all the sessions in the week, including information on the session format from future sessions in the week (eg, information on duration), allowing the checking of activity counts from previous days on the MyActivity screen, integrating the option to manually add activities for previous days, and considering how to share their data with the health care professional.

Finally, participants also said that all functionalities and guided sessions made them think that a lot of work had been put into the app and that they believed it was safe and was going to support them, which created a feeling of trust:

> *[About the guided sessions] I liked how often she came back in, because when your phone is in your pocket, sometimes it is like, "Oh God, is it still working, or has my phone shut down the app, and is it still tracking me?," but she comes in so often that it is reassuring—you don't have to keep taking your phone out to make sure that it is working.* [P3]
>
> *I really liked that it asked you how you felt at the end of it. I really like at the start the way it asks you how you are feeling and if there is [any] bad weather and stuff like that—I think that was really good. You feel safe while you are using it—you feel like it is going to be accurate. You feel like it is supporting you.* [P3]
>
> *It just feels like you have thought of everything with this, so you feel like you are in safe hands...It is nice to have the voice in the ears to motivate you.* [P3]
>
> *I think there was one day I was very tired. I suffer a bit from insomnia anyway...I was very tired when I finished. Whether it was my imagination...the session seemed to adapt to that.* [P4]

Potential Benefits

Participants indicated that the app had the potential to work well for survivors of BC as an incentive, not just because of the walks but also as it allows the users to check their progress, which they believed was motivating. It could be beneficial at different stages during the treatment, particularly in those who have never done any exercise before or who are trying to get back into it after treatment:

> *I think it would be quite high [the potential benefit], I really do. I would imagine [...] at least 60/65 percent [engaged], if not higher, [for it to] benefit.* [P1]
>
> *[...] for the first eight weeks at least after your radiotherapy [means] you're tired, you're raw, and you're scared. I think getting out with the app and getting walking at that point would be good, both physically and mentally.* [P1]
>
> *I think it could make an absolute brilliant overall meaningful exercise and moving forward positively.* [P1]
>
> *I think it would work really well, yes.* [P2]
>
> *Because it is the incentive that they have, it isn't just about going for a walk—it's going for a walk, but you can track your progress, which I think is quite motivating to say. [...] Yes, just different stages during treatment I think could be very beneficial, yes, or particularly patients where they haven't done any exercise and now, they're trying to get back into it.*

It is motivating, it is encouraging to say that it's there. [P2]

I think it would really help people. When you finish treatment, you don't know where to go, you don't know what to do, [or] what is safe to do—you question everything. I think to have something like this on your phone just for you, [to] set yourself your targets, get out, and do it—I think it would be so beneficial for so many people. [P3]

I think that the target of three a week is really good because you know that is what you have to do. I think the step tracker is so great for people who don't have a smart watch. Giving you information like that can only be beneficial to your health. [P3]

Discussion

Principal Findings

This study evaluated the potential acceptability and several aspects of user experience of mobile-based interventions for survivors of BC. A mixed methods case series study design was used to provide a deeper understanding of the individual experiences of participants. The findings cover aspects associated with the feasibility of the PA intervention, affective attitude toward the system, coherence and usability, system burden, perceived impact and quality, and potential effectiveness.

Overall, the participants found the system friendly and easy to use and showed a very positive attitude toward its various system features and aesthetics. Participants' scores on the usability and quality scales were good and seemed consistent. They perceived that the system was encouraging, increased their consciousness of their PA, and pushed them to go out more, which induced positive feelings. Furthermore, the participants had positive opinions toward the guided sessions, looking at their activity progress and history and being recognized for their achievements. All participants found it feasible and fairly easy to integrate into their daily lives and had very positive perspectives on having defined goals and being able to reschedule sessions.

This study reinforced the advantages of following user-centered design approaches and involving users at different design stages of the product. The findings from this study will inform the development of the next iteration of the system to maximize usability and the potential acceptability across a wider range of survivors of BC, which may improve the future success of the system [41]. We believe that the results of this preliminary evaluation point to the potential of the tool proposed in this thesis to support PA in survivors of BC. In this study, several challenges related to user-centered design emerged. The study has to be conducted in a way to minimize clinical disruptions, and consequently, the time for follow-up (eg, flexible duration of the intervention) was decided to minimize the burden on patients and clinicians (eg, not setting up study visits that are not aligned with the clinical visits). We also decided to provide smartphones to minimize issues with some patients using older phones with limited capacity and performance. This might also pose a challenge in the sense that some users might not have been familiarized with the provided smartphones. Overall, this preliminary feasibility study should always balance the need to be as close as possible to the real world and the potential minimization of bias in the study.

Potential Acceptability and Feasibility

Step count monitoring leads to short- and long-term step count increases [42], and in our study, participants mentioned being conscious of the step goals but not following them every day, which is reflected in lower compliance with that aspect of training in all participants. This can also be because of the need to use an extra phone for the study, which seemed to be slightly challenging, as they needed some time to get used to the different operating system and sometimes left it at home. Nevertheless, the strategy for using the extra phone was to obtain more accurate feedback on the system measurements. The code provides a basic step tracking tool implementation using the Android accelerometer signal and a basic peak detection algorithm [43] to detect when the user takes a step.

The participants seemed to agree on the feasibility of the training plan and the format of the sessions, with the different phases at a certain pace. In addition, they liked having a plan and clear targets and did not seem to feel it was overprescriptive. Participants liked the format of the sessions and that it was different for each session, and they found it positive that the sessions pushed them physically. The app adapted to the session's difficulty when patients reported feeling tired, which was perceived positively and reinforced the importance of the adaptive training functionality.

These aspects highlight the importance of involving health professionals in the design process of digital tools [44]. However, the participants had some suggestions for improvement regarding the PA plan. They argued that it would be preferable if the sessions ended with a cooling down phase. Although there are many proposed benefits of an active cooldown compared with a passive cooldown, only a few of these benefits are supported by research. However, most individuals perceived an active cooldown to be more beneficial than a passive cooldown [45], and it is important to consider it in future exercise session design. Other customizations of the PA plan, such as adding more sessions, must also be discussed individually. Nevertheless, there are specific prescription guidelines [46] that are a good practice to be followed to enhance patient safety.

Participants mentioned the challenge of keeping with the plan when it was cold or rainy. In particular, unfit adults tend not to participate in PA when the weather is unpleasant [47]. This highlights the importance of exploring ways in which the system considers and adapts to the users' context (eg, location and weather).

System usability is a critical aspect associated with intervention coherence and influences self-efficacy, which are 2 important aspects of acceptability [48]. A major usability issue revealed in the study was the poor accessibility of the weekly summary report, which was somewhat *hidden* in the history tab and, therefore, limited the participants' use of this feature. Despite these issues, the participants were confident in using the system

on their own and highlighted the importance of the initial demonstration in the first study session and of trying out the app for a while independently. The participants gave some scores close to the maximum of 5 in aesthetics and functionality, and there were some mixed scores among the 4 participants for entertainment, target group, customization, and quantity of information. Related works suggest that these aspects may be associated with system engagement [49].

From the interview data, participants seemed to agree that the app was very positive, user-friendly, useful, easy to use, and motivating. All the functionalities were perceived as very useful by at least one of the participants. The main component of the system, the guided sessions, had very positive opinions from all the participants. Commenting on the coach's communication during the session, participants suggested that having the voice with the cues about pace, progress, and encouragement while doing an activity was *lovely*. There is already solid scientific literature demonstrating the potential of digital health interventions, particularly when combining PA monitors, tailored motivational messaging, and web-based coaching, in increasing PA and having the potential to improve health outcomes [50].

The participants also reported that they felt good when they completed the sessions. Scientific research indicates that exercise is associated with positive mood changes, even when physiological benefits are not found [51]. This before and after exercise mood and fatigue feedback seems to be very important to possibly routine the adjustment of durations and intensities, which increasingly facilitates positive postexercise feelings and better maintenance of exercise [52].

With regard to personal data sharing, participants mentioned being open to sharing their personal data in exchange for a more personalized approach, which is in line with previous literature [12]. This attitude may be associated with a feeling of trust in the app, given that this was a scientific research study using an evidence-based app and validated by a health care professional (eg, oncologist and specialist nurse).

Future Considerations and Improvements

Participants identified some technical inconsistencies and provided helpful suggestions on how the mobile solution improved, which may influence the acceptability and feasibility of the system and demonstrates the importance of system evaluation with users at an early stage [53,54].

To our knowledge, there are few mobile app–based PA interventions, specifically designed for survivors of BC, that have been submitted for some type of evaluation [11,55-58]. Similarly, their findings point to the potential interest of these individuals toward a PA app, and among the successful and useful features were the balanced exercise program, visual support, viewing personal progress, activity reminders, and acknowledgment of activity achievements. However, in accordance with our findings, participants in that study perceived the exercises as being too easy to perform overall and wanted to feel more challenged. In addition, they wanted further adaptability from the app, for example, in learning from their daily routine to adjust communication. In this sense, Marcu et al [11] suggested the potential for adaptability and customization of features to increase system effectiveness. This is in line with our findings and supports the continued improvement of the system proposed in this thesis and further exploration of modules for personalized communication and adaptive PA prescription.

Overall, this study shows some promising results regarding the concept proposed for a PA coaching system for survivors of BC. Considering the aspects discussed here, an improved version of this system may have the potential to be accepted and engage these individuals, which may ultimately lead to an increase in PA adherence. Future work is required to assess and optimize the reliability of the monitoring and activity prescription systems, improve the motivational and personalization functionalities used, and assess the feasibility of the system in the long term. Only after these stages should the final step—the efficacy evaluation of the system in a rigorous trial—be considered.

Limitations

A limitation of this study is the small number of participants and also the short duration of the intervention, considering that survivors of BC do require support for long periods. This work is an early evaluation to gather preliminary insights from end users' perspectives on the concept and inform future app versions and intervention improvements. Future studies should test the acceptability and feasibility of the intervention with a higher number of participants and for longer periods before conducting more controlled trials.

The study protocol did not address the collection of data regarding recruitment (eg, the number of patients who received and opened the email with the invitation and the response rate of invitation by phone calls). This information may be highly relevant for the designing of larger studies.

All 4 participants had high levels of education and digital literacy and had at least some experience using mobile apps. In addition, all were from the same private hospital, which may be associated with an affluent socioeconomic background, better care, high education, and high awareness of the importance of self-management (eg, PA). Furthermore, the participants were moderately active and between 1 and 3 years after the main treatment. Therefore, the sample may not be representative of the wider population. Future work should consider a larger and more diverse sample of participants considering, for example, digital literacy, PA awareness and levels, the type of care, and the number of years since treatment.

Overall, PA is important for patients with BC; however, it also includes the use of resistance exercise, which was not included in this digital intervention. Future work and research in this area should also include resistance exercises, and we should consider that the sample of patients that participated in this study might not be generalizable to the general population. Our participants were moderately active and highly educated, and future research should explore how transferable the results can be to other cohorts of patients.

Conclusions

The participants of this study showed high usability and satisfaction. Further research should look into larger and more diverse cohorts and may have the potential to be acceptable and feasible for survivors of BC, particularly those in the early stages of survivorship. However, functional improvements and additional coaching content (eg, other activity types) should be considered in future iterations of the concept to be appropriate for a wider population of survivors of BC. Furthermore, more work is needed to expand system customization and automatic personalization to provide content adjusted to their individual needs and preferences at each stage in their survivorship journey. Overall, the early results presented in this study point to the potential of this tool concept to provide a friendly and satisfying coaching experience to the users, which may help improve PA adherence in survivors of BC. Therefore, this study supports future work on improving and evaluating the proposed system. Following the resolution of the technical issues experienced in this study, future evaluations of the system are needed to assess system acceptability and feasibility with a larger and more varied sample and for more prolonged periods to evaluate the system's potential for engagement and assess the influence of motivational and personalization strategies in sustaining system adherence.

Acknowledgments

This study received funding from the European Union's Horizon 2020 Research and Innovation Program: Marie Sklodowska-Curie Actions grant agreement number 722012-CATCH-Cancer: Activating Technology for Connected Health. This study was performed in collaboration with Adhera Health Inc, the University of Seville, the Insight Centre for Data Analytics from the University College Dublin, and the Beacon Hospital in Dublin. The authors would like to thank all the Beacon Hospital personnel, Professor Brian Caulfield from the Insight Centre, and the patients who participated in the study for their involvement and interest in this research.

Authors' Contributions

FMG led the study, contributed to the analysis, and wrote the main body of the manuscript. LFL, GRS, ORR, and FJNB contributed to the analysis and supported the manuscript. The authors FMG and LFL together created the interview guide.

Conflicts of Interest

GRS, FMG, FJNB, and LFL worked for at least some period of the study at Adhera Health, Inc, a digital therapeutics company that develops and commercializes mobile health solutions to support patients with cancer.

References

1. Stewart BW, Wild CP. Transitions in human development and the global cancer burden. In: Stewart BW, Wild CP, editors. World Cancer Report 2014. Lyon, France: International Agency for Research on Cancer; 2014:54-68.
2. Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. CA Cancer J Clin 2005;55(2):74-108 [FREE Full text] [doi: 10.3322/canjclin.55.2.74] [Medline: 15761078]
3. Sun YS, Zhao Z, Yang ZN, Xu F, Lu HJ, Zhu ZY, et al. Risk factors and preventions of breast cancer. Int J Biol Sci 2017 Nov 1;13(11):1387-1397 [FREE Full text] [doi: 10.7150/ijbs.21635] [Medline: 29209143]
4. Weingart SN, Brown E, Bach PB, Eng K, Johnson SA, Kuzel TM, et al. NCCN task force report: oral chemotherapy. J Natl Compr Canc Netw 2008 Mar;6 Suppl 3:S1-14. [Medline: 18377852]
5. Giese-Davis J, Collie K, Rancourt KM, Neri E, Kraemer HC, Spiegel D. Decrease in depression symptoms is associated with longer survival in patients with metastatic breast cancer: a secondary analysis. J Clin Oncol 2011 Feb 01;29(4):413-420 [FREE Full text] [doi: 10.1200/JCO.2010.28.4455] [Medline: 21149651]
6. Stein KD, Syrjala KL, Andrykowski MA. Physical and psychological long-term and late effects of cancer. Cancer 2008 Jun 01;112(11 Suppl):2577-2592. [doi: 10.1002/cncr.23448] [Medline: 18428205]
7. Bolton G, Isaacs A. Women's experiences of cancer-related cognitive impairment, its impact on daily life and care received for it following treatment for breast cancer. Psychol Health Med 2018 Dec;23(10):1261-1274. [doi: 10.1080/13548506.2018.1500023] [Medline: 30048158]
8. Spei ME, Samoli E, Bravi F, La Vecchia C, Bamia C, Benetou V. Physical activity in breast cancer survivors: a systematic review and meta-analysis on overall and breast cancer survival. Breast 2019 Apr;44:144-152. [doi: 10.1016/j.breast.2019.02.001] [Medline: 30780085]
9. Mountford N, Dorronzoro Zubiete E, Kessie T, Garcia-Zapirain B, Nuño-Solinís R, Coyle D, et al. Activating technology for connected health in cancer: protocol for a research and training program. JMIR Res Protoc 2018 Jan 24;7(1):e14 [FREE Full text] [doi: 10.2196/resprot.8900] [Medline: 29367184]
10. Palesh O, Kamen C, Sharp S, Golden A, Neri E, Spiegel D, et al. Physical activity and survival in women with advanced breast cancer. Cancer Nurs 2018;41(4):E31-E38 [FREE Full text] [doi: 10.1097/NCC.0000000000000525] [Medline: 28727578]

11. Marcu G, Misra A, Caro K, Plank M, Leader A, Barsevick A. Bounce: designing a physical activity intervention for breast cancer survivors. In: Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare. 2018 Presented at: PervasiveHealth '18; May 21-24, 2018; New York, NY, USA p. 25-34 URL: https://doi.org/10.1145/3240925.3240927 [doi: 10.1145/3240925.3240927]
12. Monteiro-Guerra F, Signorelli GR, Rivera-Romero O, Dorronzoro-Zubiete E, Caulfield B. Breast cancer survivors' perspectives on motivational and personalization strategies in mobile app-based physical activity coaching interventions: qualitative study. JMIR Mhealth Uhealth 2020 Sep 21;8(9):e18867 [FREE Full text] [doi: 10.2196/18867] [Medline: 32955446]
13. Brunet J, Taran S, Burke S, Sabiston CM. A qualitative exploration of barriers and motivators to physical activity participation in women treated for breast cancer. Disabil Rehabil 2013;35(24):2038-2045. [doi: 10.3109/09638288.2013.802378] [Medline: 23772995]
14. Smith SG, Chagpar AB. Adherence to physical activity guidelines in breast cancer survivors. Am Surg 2010 Sep;76(9):962-965. [Medline: 20836343]
15. Bluethmann SM, Bartholomew LK, Murphy CC, Vernon SW. Use of theory in behavior change interventions. Health Educ Behav 2017 Apr;44(2):245-253 [FREE Full text] [doi: 10.1177/1090198116647712] [Medline: 27226430]
16. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Adv Behav Res Ther 1978 Jan;1(4):139-161. [doi: 10.1016/0146-6402(78)90002-4]
17. Prochaska JO, DiClemente CC. Stages and processes of self-change of smoking: toward an integrative model of change. J Consult Clin Psychol 1983 Jun;51(3):390-395. [doi: 10.1037//0022-006x.51.3.390] [Medline: 6863699]
18. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]
19. Muntaner A, Vidal-Conti J, Palou P. Increasing physical activity through mobile device interventions: a systematic review. Health Informatics J 2016 Sep;22(3):451-469 [FREE Full text] [doi: 10.1177/1460458214567004] [Medline: 25649783]
20. Gosling SD, Mason W. Internet research in psychology. Annu Rev Psychol 2015 Jan 03;66:877-902. [doi: 10.1146/annurev-psych-010814-015321] [Medline: 25251483]
21. Bort-Roig J, Gilson ND, Puig-Ribera A, Contreras RS, Trost SG. Measuring and influencing physical activity with smartphone technology: a systematic review. Sports Med 2014 May;44(5):671-686. [doi: 10.1007/s40279-014-0142-5] [Medline: 24497157]
22. Harari GM, Lane ND, Wang R, Crosier BS, Campbell AT, Gosling SD. Using smartphones to collect behavioral data in psychological science: opportunities, practical considerations, and challenges. Perspect Psychol Sci 2016 Nov;11(6):838-854 [FREE Full text] [doi: 10.1177/1745691616650285] [Medline: 27899727]
23. Dorsey ER, Yvonne Chan YF, McConnell MV, Shaw SY, Trister AD, Friend SH. The use of smartphones for health research. Acad Med 2017 Feb;92(2):157-160. [doi: 10.1097/ACM.0000000000001205] [Medline: 27119325]
24. Bender JL, Yue RY, To MJ, Deacken L, Jadad AR. A lot of action, but not in the right direction: systematic review and content analysis of smartphone applications for the prevention, detection, and management of cancer. J Med Internet Res 2013 Dec 23;15(12):e287 [FREE Full text] [doi: 10.2196/jmir.2661] [Medline: 24366061]
25. Mosa AS, Yoo I, Sheets L. A systematic review of healthcare applications for smartphones. BMC Med Inform Decis Mak 2012 Jul 10;12:67 [FREE Full text] [doi: 10.1186/1472-6947-12-67] [Medline: 22781312]
26. Wang J, Wang Y, Wei C, Yao NA, Yuan A, Shan Y, et al. Smartphone interventions for long-term health management of chronic diseases: an integrative review. Telemed J E Health 2014 Jun;20(6):570-583. [doi: 10.1089/tmj.2013.0243] [Medline: 24787747]
27. Phillips SM, Courneya KS, Welch WA, Gavin KL, Cottrell A, Nielsen A, et al. Breast cancer survivors' preferences for mHealth physical activity interventions: findings from a mixed methods study. J Cancer Surviv 2019 Apr;13(2):292-305 [FREE Full text] [doi: 10.1007/s11764-019-00751-3] [Medline: 30912011]
28. De Vito Dabbs A, Myers BA, Mc Curry KR, Dunbar-Jacob J, Hawkins RP, Begey A, et al. User-centered design and interactive health technologies for patients. Comput Inform Nurs 2009;27(3):175-183 [FREE Full text] [doi: 10.1097/NCN.0b013e31819f7c7c] [Medline: 19411947]
29. Monteiro-Guerra F, Rivera-Romero O, Fernandez-Luque L, Caulfield B. Personalization in real-time physical activity coaching using mobile applications: a scoping review. IEEE J Biomed Health Inform 2020 Jun;24(6):1738-1751. [doi: 10.1109/JBHI.2019.2947243] [Medline: 31751254]
30. Jarrahi MH, Gafinowitz N, Shin G. Activity trackers, prior motivation, and perceived informational and motivational affordances. Pers Ubiquit Comput 2017 Nov 18;22(2):433-448. [doi: 10.1007/s00779-017-1099-9]
31. Epstein DA, Caraway M, Johnston C, Ping A, Fogarty J, Munson SA. Beyond abandonment to next steps: understanding and designing for life after personal informatics tool use. Proc SIGCHI Conf Hum Factor Comput Syst 2016 May;2016:1109-1113 [FREE Full text] [doi: 10.1145/2858036.2858045] [Medline: 28503678]
32. Sekhon M, Cartwright M, Francis JJ. Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. BMC Health Serv Res 2017 Jan 26;17(1):88 [FREE Full text] [doi: 10.1186/s12913-017-2031-8] [Medline: 28126032]

33. Arthanat S, Desmarais JM, Eikelberg P. Consumer perspectives on the usability and value of the iBOT(®) wheelchair: findings from a case series. Disabil Rehabil Assist Technol 2012 Mar;7(2):153-167. [doi: 10.3109/17483107.2011.589487] [Medline: 21967371]
34. Paul L, Brewster S, Wyke S, McFadyen AK, Sattar N, Gill JM, et al. Increasing physical activity in older adults using STARFISH, an interactive smartphone application (app); a pilot study. J Rehabil Assist Technol Eng 2017 Apr 5;4:2055668317696236 [FREE Full text] [doi: 10.1177/2055668317696236] [Medline: 31186925]
35. Diener E, Emmons RA, Larsen RJ, Griffin S. The satisfaction with Life Scale. J Pers Assess 1985 Feb;49(1):71-75. [doi: 10.1207/s15327752jpa4901_13] [Medline: 16367493]
36. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003 Aug;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]
37. Frøkjær E, Hertzum M, Hornbæk K. Measuring usability: are effectiveness, efficiency, and satisfaction really correlated? In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2000 Presented at: CHI '00; April 1-6, 2000; The Hague, The Netherlands p. 345-352 URL: https://orbit.dtu.dk/en/publications/measuring-usability-are-effectiveness-efficiency-and-satisfaction [doi: 10.1145/332040.332455]
38. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016 Jun 10;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]
39. User Experience Design. Semantic Studios. 2004. URL: https://semanticstudios.com/user_experience_design/ [accessed 2021-06-14]
40. Monteiro-Guerra F, Signorelli GR, Tadas S, Dorronzoro Zubiete E, Rivera Romero O, Fernandez-Luque L, et al. A personalized physical activity coaching app for breast cancer survivors: design process and early prototype testing. JMIR Mhealth Uhealth 2020 Jul 15;8(7):e17552 [FREE Full text] [doi: 10.2196/17552] [Medline: 32673271]
41. Zapata BC, Fernández-Alemán JL, Idri A, Toval A. Empirical studies on usability of mHealth apps: a systematic literature review. J Med Syst 2015 Feb;39(2):1. [doi: 10.1007/s10916-014-0182-2] [Medline: 25600193]
42. Chaudhry UA, Wahlich C, Fortescue R, Cook DG, Knightly R, Harris T. The effects of step-count monitoring interventions on physical activity: systematic review and meta-analysis of community-based randomised controlled trials in adults. Int J Behav Nutr Phys Act 2020 Oct 09;17(1):129 [FREE Full text] [doi: 10.1186/s12966-020-01020-8] [Medline: 33036635]
43. Mladenov M, Mock M. A step counter service for Java-enabled devices using a built-in accelerometer. In: Proceedings of the 1st International Workshop on Context-Aware Middleware and Services: affiliated with the 4th International Conference on Communication System Software and Middleware (COMSWARE 2009). 2009 Presented at: CAMS '09; June 16, 2009; Dublin, Ireland p. 1-5. [doi: 10.1145/1554233.1554235]
44. Giunti G, Giunta DH, Guisado-Fernandez E, Bender J, Fernandez-Luque L. A biopsy of breast cancer mobile applications: state of the practice review. Int J Med Inform 2018 Feb;110:1-9 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.10.022] [Medline: 29331247]
45. Van Hooren B, Peake JM. Do we need a cool-down after exercise? A narrative review of the psychophysiological effects and the effects on performance, injuries and the long-term adaptive response. Sports Med 2018 Jul;48(7):1575-1595 [FREE Full text] [doi: 10.1007/s40279-018-0916-2] [Medline: 29663142]
46. Rock CL, Doyle C, Demark-Wahnefried W, Meyerhardt J, Courneya KS, Schwartz AL, et al. Nutrition and physical activity guidelines for cancer survivors. CA Cancer J Clin 2012;62(4):243-274 [FREE Full text] [doi: 10.3322/caac.21142] [Medline: 22539238]
47. Aspvik NP, Viken H, Ingebrigtsen JE, Zisko N, Mehus I, Wisløff U, et al. Do weather changes influence physical activity level among older adults? - the Generation 100 study. PLoS One 2018 Jul 6;13(7):e0199463 [FREE Full text] [doi: 10.1371/journal.pone.0199463] [Medline: 29979711]
48. Chiu TM, Eysenbach G. Stages of use: consideration, initiation, utilization, and outcomes of an Internet-mediated intervention. BMC Med Inform Decis Mak 2010 Nov 23;10:73 [FREE Full text] [doi: 10.1186/1472-6947-10-73] [Medline: 21092275]
49. Vandelanotte C, Müller AM, Short CE, Hingle M, Nathan N, Williams SL, et al. Past, present, and future of eHealth and mHealth research to improve physical activity and dietary behaviors. J Nutr Educ Behav 2016 Mar;48(3):219-28.e1. [doi: 10.1016/j.jneb.2015.12.006] [Medline: 26965100]
50. Davies CA, Spence JC, Vandelanotte C, Caperchione CM, Mummery WK. Meta-analysis of Internet-delivered interventions to increase physical activity levels. Int J Behav Nutr Phys Act 2012 Apr 30;9:52 [FREE Full text] [doi: 10.1186/1479-5868-9-52] [Medline: 22546283]
51. Annesi JJ. Effects of minimal exercise and cognitive behavior modification on adherence, emotion change, self-image, and physical change in obese women. Percept Mot Skills 2000 Aug;91(1):322-336. [doi: 10.2466/pms.2000.91.1.322] [Medline: 11011903]
52. Annesi JJ. Relationship between before-to-after-exercise feeling state changes and exercise session attendance over 14 weeks: testing principles of Operant Conditioning. Eur J Sport Sci 2005 Dec;5(4):159-163. [doi: 10.1080/17461390500387056]

53. Lenarduzzi V, Taibi D. MVP explained: a systematic mapping study on the definitions of minimal viable product. In: Proceedings of the 42th Euromicro Conference on Software Engineering and Advanced Applications. 2016 Presented at: SEAA '16; August 31-September 2, 2016; Limassol, Cyprus p. 112-119. [doi: 10.1109/seaa.2016.56]
54. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016 Nov;51(5):843-851 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
55. Lozano-Lozano M, Galiano-Castillo N, Martín-Martín L, Pace-Bedetti N, Fernández-Lao C, Arroyo-Morales M, et al. Monitoring energy balance in breast cancer survivors using a mobile app: reliability study. JMIR Mhealth Uhealth 2018 Mar 27;6(3):e67 [FREE Full text] [doi: 10.2196/mhealth.9669] [Medline: 29588273]
56. Phillips SM, Collins LM, Penedo FJ, Courneya KS, Welch W, Cottrell A, et al. Optimization of a technology-supported physical activity intervention for breast cancer survivors: Fit2Thrive study protocol. Contemp Clin Trials 2018 Mar;66:9-19 [FREE Full text] [doi: 10.1016/j.cct.2018.01.001] [Medline: 29330081]
57. Chung IY, Jung M, Park YR, Cho D, Chung H, Min YH, et al. Exercise promotion and distress reduction using a mobile app-based community in breast cancer survivors. Front Oncol 2020 Jan 10;9:1505 [FREE Full text] [doi: 10.3389/fonc.2019.01505] [Medline: 31998651]
58. Phillips S, Solk P, Welch W, Auster-Gussman L, Lu M, Cullather E, et al. A technology-based physical activity intervention for patients with metastatic breast cancer (Fit2ThriveMB): protocol for a randomized controlled trial. JMIR Res Protoc 2021 Apr 23;10(4):e24254 [FREE Full text] [doi: 10.2196/24254] [Medline: 33890857]

Abbreviations

BC: breast cancer
BCT: behavioral change theory
mHealth: mobile health
PA: physical activity
QoL: quality of life
TFA: theoretical framework of acceptability

Edited by A Mavragani; submitted 23.07.21; peer-reviewed by A González-Santos, M Lopez Garzon, K Wendrich; comments to author 21.10.21; revised version received 03.04.22; accepted 16.05.22; published 22.06.22.

Please cite as:
Signorelli GR, Monteiro-Guerra F, Rivera-Romero O, Núñez-Benjumea FJ, Fernández-Luque L
Breast Cancer Physical Activity Mobile Intervention: Early Findings From a User Experience and Acceptability Mixed Methods Study
JMIR Form Res 2022;6(6):e32354
URL: https://formative.jmir.org/2022/6/e32354
doi:10.2196/32354
PMID:35731554

Original Paper

A Web-Based mHealth Intervention With Telephone Support to Increase Physical Activity Among Pregnant Patients With Overweight or Obesity: Feasibility Randomized Controlled Trial

Tainayah Thomas[1], PhD; Fei Xu[2], MSc; Sneha Sridhar[2], MPH; Tali Sedgwick[2], RDN; Linda Nkemere[2], BA; Sylvia E Badon[2], PhD; Charles Quesenberry[2], PhD; Assiamira Ferrara[2], MD, PhD; Sarah Mandel[3], MD; Susan D Brown[2,4,5], PhD; Monique Hedderson[2], PhD

[1]Department of Epidemiology and Population Health, Stanford University School of Medicine, Palo Alto, CA, United States

[2]Division of Research, Kaiser Permanente Northern California, Oakland, CA, United States

[3]The Permanente Medical Group, San Francisco, CA, United States

[4]School of Medicine, University of California, Davis, Sacramento, CA, United States

[5]Kaiser Permanente Bernard J. Tyson School of Medicine, Pasadena, CA, United States

Corresponding Author:
Tainayah Thomas, PhD
Department of Epidemiology and Population Health
Stanford University School of Medicine
1701 Page Mill Road
Palo Alto, CA, 94304
United States
Phone: 1 3107211208
Email: tainayah@stanford.edu

Abstract

Background: Pregnant patients with overweight or obesity are at high risk for perinatal complications. Excess gestational weight gain (GWG) further exacerbates this risk. Mobile health (mHealth) lifestyle interventions that leverage technology to facilitate self-monitoring and provide just-in-time feedback may motivate behavior change to reduce excess GWG, reduce intervention costs, and increase scalability by improving access.

Objective: This study aimed to test the acceptability and feasibility of a pilot mHealth lifestyle intervention for pregnant patients with overweight or obesity to promote moderate intensity physical activity (PA), encourage guideline-concordant GWG, and inform the design of a larger pragmatic cluster randomized controlled trial.

Methods: We conducted a mixed methods acceptability and feasibility randomized controlled trial among pregnant patients with a prepregnancy BMI of 25 to 40 kg/m^2. Patients with singletons at 8 to 15 weeks of gestation who were aged ≥21 years and had Wi-Fi access were recruited via email from 2 clinics within Kaiser Permanente Northern California and randomized to receive usual prenatal care or an mHealth lifestyle intervention. Participants in the intervention arm received wireless scales, access to an intervention website, activity trackers to receive automated feedback on weight gain and activity goals, and monthly calls from a lifestyle coach. Surveys and focus groups with intervention participants assessed intervention satisfaction and ways to improve the intervention. PA outcomes were self-assessed using the Pregnancy Physical Activity Questionnaire, and GWG was assessed using electronic health record data for both arms.

Results: Overall, 33 patients were randomly assigned to the intervention arm, and 35 patients were randomly assigned to the usual care arm. All participants in the intervention arm weighed themselves at least once a week, compared with 20% (7/35) of the participants in the usual care arm. Participants in the intervention arm wore the activity tracker 6.4 days per week and weighed themselves 5.3 times per week, and 88% (29/33) of them rated the program “good to excellent.” Focus groups found that participants desired more nutrition-related support to help them manage GWG and would have preferred an app instead of a website. Participants in the intervention arm had a 23.46 metabolic equivalent of task hours greater change in total PA per week and a 247.2-minute greater change in moderate intensity PA per week in unadjusted models, but these effects were attenuated in adjusted models (change in total PA: 15.55 metabolic equivalent of task hours per week; change in moderate intensity PA: 199.6 minutes per week). We found no difference in total GWG (mean difference 1.14 kg) compared with usual care.

Conclusions: The pilot mHealth lifestyle intervention was feasible, highly acceptable, and promoted self-monitoring. Refined interventions are needed to effectively affect PA and GWG among pregnant patients with overweight or obesity.

Trial Registration: ClinicalTrials.gov NCT03936283; https://clinicaltrials.gov/ct2/show/NCT03936283

(JMIR Form Res 2022;6(6):e33929) doi:10.2196/33929

KEYWORDS

mobile health; gestational weight gain; obesity; physical activity; mobile phone

Introduction

Background

Pregnant patients with overweight or obesity (BMI ≥25 kg/m^2) are at high risk for perinatal complications, including gestational diabetes, pre-eclampsia, excess fetal growth, birth injuries, and cesarean section [1,2]. Excess gestational weight gain (GWG) further exacerbates the elevated risk of perinatal complications in pregnant patients with overweight or obesity. More than half of pregnant patients with overweight or obesity exceed the recommended amount of GWG [3,4].

Pregnancy is a unique window in which patients are often motivated to make healthy changes, which presents an unparalleled opportunity to intervene in health behaviors. Intensive behavioral interventions requiring in-person counseling with multiple clinic visits may not be feasible for many patients. Technology such as mobile health (mHealth) interventions can deliver automated, standardized information that eliminates social barriers [5-7], while potentially reducing intervention costs [8-12] and improving quality [10,13-15].

Usual prenatal care includes regular weight measurements at prenatal care visits; however, during early pregnancy, these visits are infrequent, and patients may gain more weight between visits than that suggested by the Institute of Medicine (IOM) GWG guidelines. This makes it critical to assess the impact of promoting self-weighing between visits and sharing this information with the patients' clinical care team to better evaluate and support guideline-concordant weight gain.

Research has shown that physical activity (PA) alone can reduce GWG [16-19]. For example, a meta-analysis of 12 intervention trials assessing the association between PA during pregnancy and GWG found a significantly lower average GWG in the intervention group than in the control group [19]. Current guidelines recommend pregnant patients get ≥30 minutes per day of moderate intensity PA, most days of the week [20], but this goal is rarely achieved [21]. Therefore, interventions to improve PA, especially among pregnant patients with overweight and obesity are needed. Commercially available technologies such as activity trackers (shown to increase PA in nonpregnant adults [22]) and wireless scales enable real-time self-monitoring, goal setting, and tailored feedback on goals. Tailored feedback has been successful in reinforcing motivation for behavior change, especially when delivered in relation to goal attainment [23-25]. Using technology to facilitate behavior change has the advantage of providing a resource that patients can use conveniently without disrupting their busy lives. In addition, self-monitoring of weight with wireless scales can transmit weight data directly to health coaches and clinicians, allow for objective measurement of daily weight, increase adherence compared with paper monitoring [26], and promote weight loss in adults with overweight BMI when used in conjunction with additional behavior change techniques [27]. Although wireless scales and activity trackers may facilitate self-monitoring, provide just-in-time feedback, and motivate change, few studies have both quantitatively and qualitatively evaluated the feasibility, acceptability, and efficacy of technology-based mHealth lifestyle interventions in pregnant patients with overweight or obesity.

Objectives

This pilot acceptability and feasibility randomized controlled trial was developed to inform the design of a larger pragmatic cluster randomized controlled trial. The objective of this mixed methods study was to test the acceptability and feasibility of a pilot mHealth lifestyle intervention for pregnant patients with overweight or obesity to promote self-monitoring of weight and PA and encourage guideline-concordant GWG. As such, the primary aim was to investigate whether it was feasible to implement the intervention with the target population. The study also assessed the perceived usefulness of the intervention to determine whether it was acceptable to the study participants. Finally, the pilot acceptability and feasibility randomized controlled trial explored the preliminary efficacy findings (ie, PA and GWG) using adjusted intention-to-treat analyses.

Methods

Study Design

Study of a Randomized Intervention Designed to Increase Exercise in Pregnancy (STRIDE) is a 2-arm, parallel group pilot randomized controlled trial (ClinicalTrials.gov identifier: NCT03936283) conducted between May 2017 and May 2018. We used a mixed methods study design to assess the acceptability and feasibility of the STRIDE mHealth lifestyle intervention in a sample of pregnant patients with overweight or obesity from 2 Kaiser Permanente Northern California (KPNC) medical clinics. Participants were randomized to receive usual care (n=35) or usual care plus an mHealth lifestyle intervention (n=33). Participants completed a web-based survey at 10 and 33 weeks of gestation. To better understand the perspectives of intervention participants, participants in the intervention arm completed a program evaluation survey (n=33) and participated in 3 focus groups (n=14).

Ethics Approval

STRIDE was approved by the KPNC Institutional Review Board (approval number 1278778).

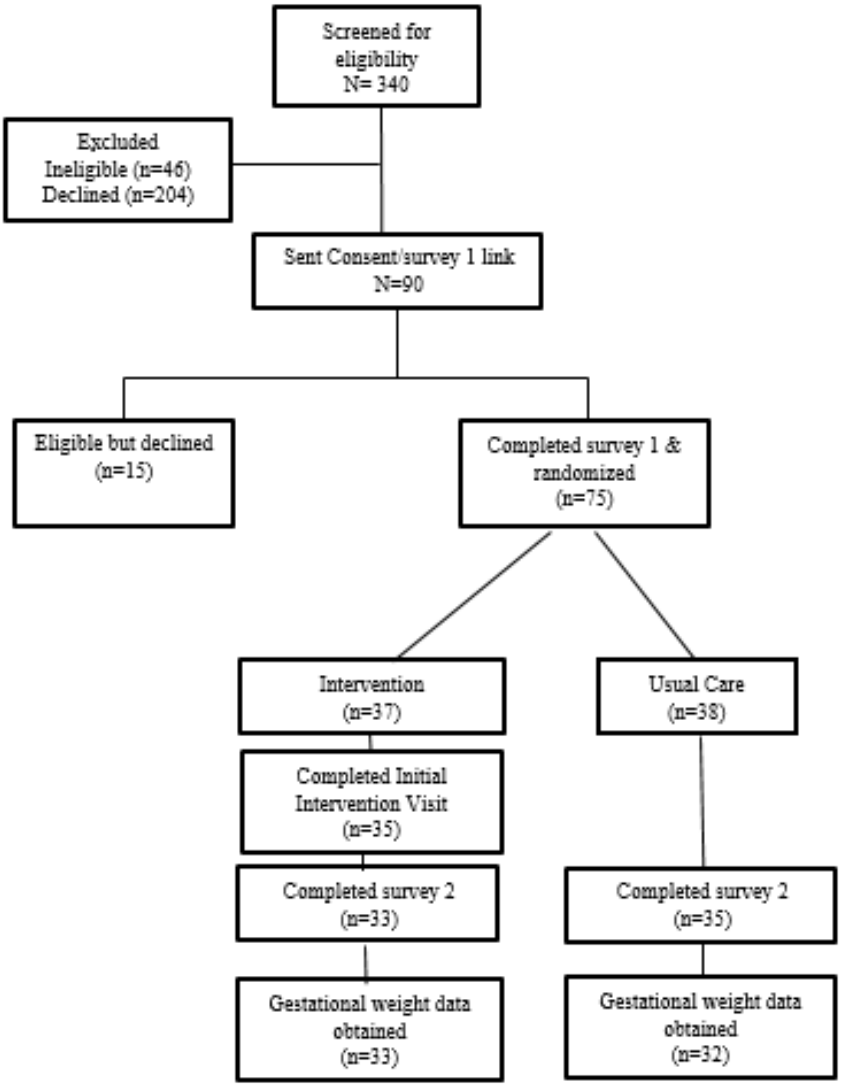
Eligibility

Pregnant patients who were at <12 gestational weeks and received care at KPNC medical centers were first identified in the electronic health record (EHR). Eligibility criteria were as follows: (1) aged ≥21 years; (2) prepregnancy BMI between 25 and 40 kg/m^2 (based on weight measured in the clinical setting within 12 months before the last menstrual period, or if unavailable, the first weight measured within the first 10 weeks of pregnancy); and (3) a singleton pregnancy. Eligibility was further assessed through a tiered process beginning with approval from medical providers to contact each patient and EHR review. Medical exclusion criteria that may affect outcome assessment, evaluated by the EHR review and interview during a recruitment screening call, included multiple gestation, pregnancy loss, high-risk pregnancy (ie, drug or alcohol abuse, chronic health problems, or pregnancy complications), thyroid disease diagnosed in the last 30 days, and use of glucose-lowering medications or corticosteroids. Exclusion criteria that may interfere with full participation in the trial were assessed starting at the recruitment call and included plans to move out of the area or change health plan membership during pregnancy, no reliable access to a smartphone and Wi-Fi at home, inability to communicate in English, and unwillingness to be randomized.

Randomization, Recruitment, and Masking

Patients were randomly assigned to the mHealth lifestyle intervention arm or the usual care control arm upon completion of a consent form and survey 1 (Figure 1). The adaptive randomization procedure ensured that equivalent numbers of patients were assigned to each study arm and that the 2 study arms remained balanced overall and at each level of key characteristics: age (21-29.9, 30-34.9, and ≥35 years), prepregnancy BMI (25.0-29.9, 30.0-34.9, and 35.0-39.9 kg/m^2), and race and ethnicity (Asian or Pacific Islander, Black, Hispanic, White, and multiethnic or other or unknown). Study participants were recruited via email by the study staff with a follow-up phone call 1 week later. The biostatistician, clinicians, and research assistants (LN and Socorro Dalton) who sent out study-related emails and surveys and investigators were masked to study arm assignment.

Figure 1. Flowchart of the Study of a Randomized Intervention Designed to Increase Exercise in Pregnancy (STRIDE).



Usual Care

Participants were randomized to the usual care arm and received standard KPNC prenatal medical care. This includes an initial prenatal visit at 7 to 10 weeks of gestation and a newsletter containing the IOM GWG guidelines and advice on healthy eating. Participants with routine pregnancies received an additional 7 prenatal visits between 16 weeks of gestation and delivery. Medical staff weighed the patients at each visit per the standard care.

Intervention

In addition to the aforementioned usual care, patients randomized to the intervention arm received a multicomponent mHealth lifestyle intervention. The intervention targeted behavior changes for PA and weight management to help patients gain within the IOM recommended range for GWG according to their prepregnancy BMI category (7-11.5 kg for women with overweight and 5-9 kg for women with obesity [28]). Our mHealth pilot intervention was adapted from the Gestational Weight Gain and Optimal Wellness (GLOW) trial, a theory-based behavioral intervention that adapted the National Diabetes Prevention Program [29], and was delivered primarily by telehealth, for pregnant patients with overweight or obesity with the goal of reducing excess GWG [30]. The GLOW intervention consisted of 2 in-person and 11 telephone sessions on behavioral strategies to improve weight management, PA, diet, and stress management in addition to usual antenatal care. Compared with usual care only, the GLOW intervention substantially reduced the proportion of participants exceeding the IOM guidelines for weekly rate of GWG and reduced total caloric intake, proportion of calories from saturated fat, sedentary behaviors, serum leptin concentration, and markers of insulin resistance among intervention participants [30]. This pilot mHealth intervention aimed to build upon the GLOW trial and incorporate its successful components into a mobile modality.

Conceptual Framework for the Intervention

We followed a tailored, trimester-specific approach to behavior change using constructs from social cognitive theory by Bandura [31-33] and the transtheoretical model [34], which have been the basis of adherence to healthy diet and PA in past research [30,35-37]. Key components included in the mHealth tool were as follows: (1) self-monitoring: weight self-monitoring enhances weight management [38,39] and (2) goal setting: participants were encouraged to set sequential, realistic, and short-term PA goals.

Behaviors Targeted by the Intervention

PA Goals

Participants in the intervention arm were asked to set PA goals and gradually increase their activity to ultimately reach 150 minutes of moderate to vigorous activity per week, in accordance with the current American College of Obstetricians and Gynecologists recommendations [20]. Participants were provided with a Withings Activité Pop PA tracker that was worn on the wrist and tracked daily steps and minutes of moderate to vigorous intensity PA (MVPA) based on a 3-axis accelerometer optoelectronics sensor with intensity based on a metabolic coefficient ≥3. Withings activity trackers are among the most accurate for measuring steps and MVPA [40,41], with the lowest rate of false positive steps [42]. The participants were encouraged to wear their tracker daily.

Self-weighing

Participants were provided with a Withings Body digital scale that transmits weights via Wi-Fi or Bluetooth and has a margin of error on weight <200 g against a gold standard scale [43]. The participants were encouraged to weigh themselves daily at home.

Healthy Eating

Patients who desired a tool to track their diet were referred to popular free mHealth apps and websites such as MyFitnessPal and Choose MyPlate.

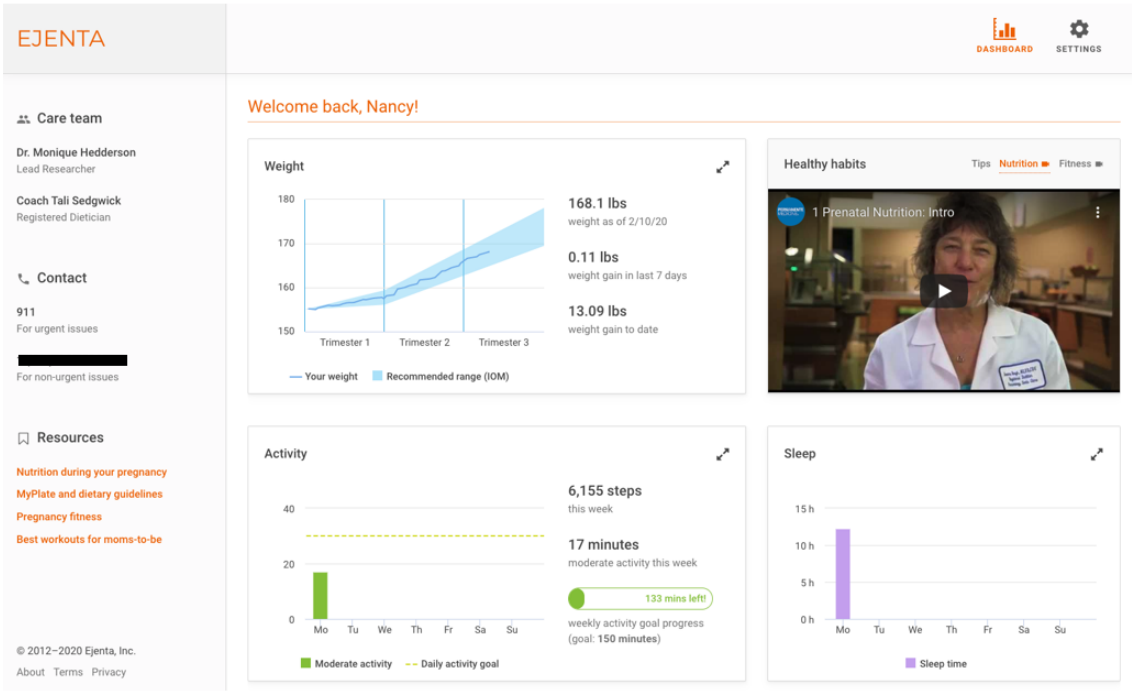
Intervention Components

The components of the intervention were as follows:

1. *Goal setting and check-in calls*—a one-time baseline visit conducted by a research assistant was held at the participant's residence. During this visit, participants were oriented to the intervention and its tools and were asked to set a baseline goal for how many minutes of MVPA they would complete during the following 7 days. Participants were given print materials including a guidebook with 5 core sessions on the following topics: (1) Welcome to the STRIDE Program, (2) Getting Started with Physical Activity, (3) Getting Started with Healthy Eating, (4) Exercise Your Options, and (5) Talk Back Negative Thoughts. A lifestyle coach, a registered dietitian nutritionist with training in motivational interviewing, performed a check-in call 1 week later to help participants evaluate progress toward their goal and set a new activity goal for the following week. Subsequently, check-in calls were conducted monthly until the end of the pregnancy.
2. *Web-based, mHealth website*—the mHealth website was developed by the study team and engineers at the technology partner Ejenta, Inc. After the website was developed, it was beta tested with 5 pregnant patients to assess usability and understanding of the website's components. The mHealth website was accessible via iPhone or Android smartphone or desktop through a unique log-in credential for each study participant. Real-time data from the activity tracker and scale were transmitted to the mHealth website. The website included a graph of participants' GWG in relation to IOM guidelines and minutes of MVPA in relation to their goals (Figure 2) and pregnancy-related resources to help participants manage their GWG. A clinician portal enabled lifestyle coaches to view the participants' self-monitoring data to tailor calls.
3. *Messages with personalized feedback*—participants received messages via email or SMS text message based on their preferences. Message content included reminders for self-weighing and self-monitoring PA; milestones at each trimester; and progress, goals, and milestones for PA (see samples in Textbox 1). Goal-achieved messages were sent whenever a participant reached a goal. Adherence reminder messages were sent the day the participant did not have

weight or activity data (separate messages for weight vs activity). Motivational messages were sent once a week to encourage participants to reach their weekly activity goal. Activity milestone (personal best weekly, personal best daily, doubled activity and meeting goal, and activity in a row) messages were sent at the end of the day on Sunday if any milestones were achieved.

Figure 2. Representative screenshot of the Study of a Randomized Intervention Designed to Increase Exercise in Pregnancy (STRIDE) mobile health website.



Textbox 1. Types of messages sent to Study of a Randomized Intervention Designed to Increase Exercise in Pregnancy (STRIDE) intervention arm participants.

Activity goal progress

- "You had 129 minutes of moderate activity from (Monday, 5/8) to (Friday, 5/12). You need 21 more moderate minutes to reach your goal by Sunday. Let's do this!"

Activity goal reached

- "Congrats on reaching your weekly activity goal! You've already been active for 150 minutes this week."

Activity milestones

- Personal best (daily)
  - "You have a new personal best 62 minutes of moderate activity on (Friday, 5/12). Congrats!"
- Doubled daily activity goal
  - "Wow. You got over double your activity goal of 30 moderate minutes with 68 minutes on (Thursday, 5/11). You rock! Keep up the good work!"
- Met daily activity goal >3 days in a row
  - "You met your daily activity goal for 3 days in a row from (Monday, 5/8) to (Wednesday, 5/10). There's no stopping you now! Power on!"

Survey Overview

Study surveys were sent via email and were administered on the web. Survey 1 was administered at baseline and survey 2 was administered at 33 to 36 gestational age (GA) weeks. The surveys covered the self-assessed domains of pregnancy history, sleep, current PA, social support, quality of life, advice from their obstetrician or gynecologist, and demographic information. For intervention participants, survey 2 also included intervention evaluation questions. Survey participants received a US $30 Amazon gift card after completing both survey 1 and survey 2.

Intervention Evaluation Outcomes

To assess the acceptability and feasibility of this pilot intervention, we analyzed adherence to self-monitoring (eg, wireless scale and wearable tracker) data and conducted surveys and focus groups with intervention participants to explore their satisfaction and experiences with the intervention. In the survey, participants ranked each intervention component on a 4-point Likert scale from very helpful to not at all helpful. Participants also rated the intervention as a whole on a 4-point Likert scale from fair to excellent and responded to whether they would recommend the program to other pregnant women on a 3-point Likert scale from *probably not* to *definitely yes*. In the focus groups, moderator guide questions examined suggestions for intervention improvement, impediments to regular use of intervention components, and overall intervention likes and dislikes. Focus groups were moderated by a registered dietitian and lifestyle coach and were conducted via WebEx. The focus group participants received a US $50 Amazon gift card for their participation.

We set a feasibility cutoff for self-monitoring (eg, self-weighing and wearing a wearable tracker) at ≥5 days per week as feasible for the participants. We set an acceptability cutoff for the perceived usefulness of the intervention at ≥80% helpful or very helpful responses for intervention component responses and good or very good or excellent for overall intervention responses to survey questions as acceptable for participants. We analyzed the focus group data to better understand what worked well for participants and to obtain information on how to improve the intervention for the larger pragmatic trial.

Exploratory Outcomes

Although this pilot study was not powered for clinical outcomes, in addition to examining intervention acceptability and feasibility, we also assessed the intervention efficacy for various exploratory outcomes. The primary exploratory outcome was PA, measured both as total activity and in metabolic equivalent of task (MET) for that activity. MET is a measure of the intensity of PA. MET hours per week (total, moderate, or vigorous sports or exercise or moderate sports or exercise) and in activity minutes per week (moderate sports exercise). PA was self-assessed using the Pregnancy Physical Activity Questionnaire (PPAQ) [44] in both intervention arms. The PPAQ is an accurate and reliable measure of PA during pregnancy [44]. Participants reported the time spent in various PAs in the 2 months before completing the study. PAs were assessed in 5 domains: household or caregiving (13 activities), occupational (5 activities), sports and exercise (12 activities), transportation (3 activities), and inactivity or sedentary behavior (3 activities). For every activity, the participants selected 1 of 6 categorical responses for the time spent in that activity. Categorical responses included none, <0.5 hours per day, 0.5 to almost 1 hour per day, 1 to almost 2 hours per day, 2 to almost 3 hours per day, and ≥3 hours per day. Energy expended for each activity was calculated by multiplying the midpoint of the duration category reported spent in the activity by the corresponding MET for that activity. MET values for walking and light to moderate intensity household tasks were based on field-based measurements of pregnant women [45]. MET values for all other activities were based on the Compendium of Physical Activities [46]. Total duration and energy expenditure was calculated overall (including light, moderate, and vigorous PA) and separately for each intensity of PA (sedentary: <1.5 METs; light: 1.5 to <3 METs; moderate: 3 to 6 METs; vigorous: >6 METs). PA duration and energy expenditure overall and by intensity categories were the outcomes of interest. All patient weights were clinically assessed during the prenatal visits. Total GWG was defined as the last measured weight within 3 weeks before delivery minus the first measured weight after conception and up to 13 weeks of GA. The rate of total GWG was defined as the total GWG divided by the difference in GA weeks between the first and last measured weights during pregnancy. We also assessed the following perinatal outcomes as potential adverse events by using EHR data: gestational diabetes mellitus, preterm birth, large-for-GA, small-for-GA, and cesarean section.

Analysis

All quantitative statistical analyses were conducted with SAS (version 9.4; SAS Institute Inc) and performed according to randomized group assignment (intention to treat), which included all participants for whom PA survey data or pregnancy weight measured after randomization were available. Multiple linear regression was used to estimate the point estimates and CIs of the overall difference between the usual care and intervention groups in change in activity in MET hours per week, change in activity in minutes per week, and GWG. All analyses were adjusted for the variables used in the adaptive randomization procedure and prepregnancy weight for GWG outcomes and baseline PA for PA outcomes. Therefore, our adjusted model for change in PA was adjusted for PA at baseline survey, age, parity, prepregnancy BMI, race or ethnicity, and difference in GA weeks between the baseline survey and the second survey. The adjusted model for GWG was adjusted for age, parity, prepregnancy BMI, race or ethnicity, and difference in GA weeks between the last and first measured weight during pregnancy.

To analyze the intervention evaluation data, we summarized the responses to the survey questions using frequencies. The focus groups were audio recorded and transcribed. We conducted directed content analysis [47] to analyze the focus group data. A coding guide was developed a priori based on the study's conceptual framework, intervention components, and interview guide. The focus group transcripts were read to derive additional codes by highlighting words from the text that appeared to capture key thoughts or concepts. Labels for codes were developed that reflected more than one key thought. The final coding guide included codes reflecting topics from the conceptual framework and interview guide, intervention components, and inductively identified de novo topics. Codes were applied to the entire data set. Matrices were used to visually represent the data and to facilitate analysis by organizing and reducing the data.

Results

Overview

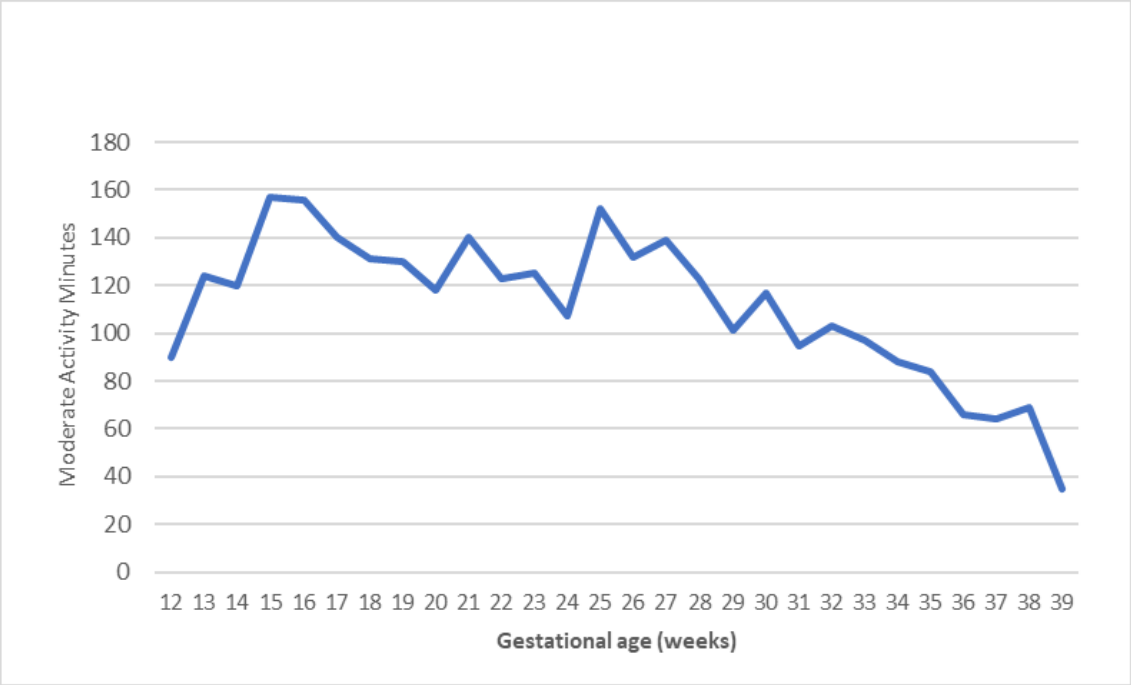
Among the 340 pregnant patients screened for eligibility, 46 (13.5%) were excluded owing to ineligibility and 204 (60%)

declined or were unable to be reached. Among the 90 eligible patients, 75 (83%) consented to randomization. Following randomization, 89% (33/37) of the intervention participants and 92% (35/38) of the usual care participants completed all study surveys (*n*=35), and 89% (33/37) of the intervention participants and 84% (32/38) of the usual care participants had a weight measurement at the end of pregnancy (Figure 1). Reasons for loss to follow-up included pregnancy loss, change to insurance, and maximum contact attempts exceeded. The 2 study arms had similar baseline characteristics (Table 1). On average, participants in the intervention condition had a high adherence to self-monitoring: they wore the tracker 6 days per week and weighed themselves 5 times per week. In addition, in late pregnancy, 100% (33/33) of women in the intervention arm reported weighing themselves at least once a week compared with 20% (7/35) of women in the usual care arm. The intervention participants had an average of 112 mean minutes (SD 30 minutes) of moderate activity per week based on wearable tracker data across the entire intervention (Figure 3). On average, the mean minutes of moderate activity per week increased early in the intervention to a maximum of 157 minutes at 15 weeks of gestation and decreased during the third trimester to a minimum of 35 minutes at 39 weeks of gestation.

Table 1. Baseline characteristics by treatment condition: the Study of a Randomized Intervention Designed to Increase Exercise in Pregnancy (STRIDE) randomized controlled trial.

	Intervention (N=33)	Usual care (N=35)	*P* value
Age (years), mean (SD)	34.8 (4.2)	33.2 (3.7)	.11
Prepregnancy BMI (kg/m^2), mean (SD)	28.9 (2.5)	28.9 (2.6)	
25.0 to 29.9, n (%)	24 (73)	26 (74)	.94
30.0 to 40.0, n (%)	9 (27)	9 (26)	.88
Race-ethnicity, n (%)			.82
Asian	5 (15)	5 (14)	
White	18 (55)	22 (63)	
Hispanic	3 (9)	4 (11)	
African American	1 (3)	1 (3)	
Multiracial or other	6 (18)	3 (9)	
Parity, n (%)			.47
0	20 (61)	17 (49)	
1	12 (36)	15 (43)	
>2	1 (3)	3 (9)	
Household income per year (US $), n (%)			.42
<100,000	4 (12)	7 (20)	
100,000 to 199,999	16 (49)	19 (54)	
≥200,000	13 (39)	9 (26)	
Education, n (%)			.61
High school or some college	3 (9)	6 (17)	
College graduate (4-year course)	12 (36)	11 (31)	
Postgraduate degree	18 (55)	18 (51)	
Gestational week at survey 1, mean (SD)	11.0 (1.8)	10.7 (1.3)	.36
Gestational week at survey 2, mean (SD)	33.6 (1.0)	33.2 (0.4)	.04

Figure 3. Mean moderate activity (minutes per week) by intervention group participants of the Study of a Randomized Intervention Designed to Increase Exercise in Pregnancy (STRIDE).



Intervention Acceptability

We conducted an evaluation survey and 3 focus groups to better understand participants' experiences with the intervention. A total of 33 participants in the intervention arm completed the evaluation survey (89% response rate). Overall, the mHealth lifestyle intervention was rated highly, with 88% (29/33) of the participants rating the intervention as excellent, very good, or good, and 85% (28/33) of the participants reporting that they would recommend the intervention to other pregnant patients.

A total of 22 participants agreed to be contacted for focus groups and were invited to participate in the qualitative study via email. A total of 14 participants enrolled in the 3 focus groups (64% response rate). The focus groups lasted for 40 to 50 minutes. Table 2 lists the ratings of each intervention component and the illustrative quotes related to the component. The digital scale was rated most highly (32/33, 97% of the participants rated it as very or moderately helpful) among all the intervention components followed by the coach calls (26/33, 79% of the participants rated it as very or moderately helpful), PA tracker, and text messages (24/33, 73% of the participants rated it as very or moderately helpful), with the mHealth website rated lowest (19/33, 58% of the participants rated it as very or moderately helpful).

Table 2. Program evaluation and acceptability results.

Intervention component	Survey results (n=33); "On the basis of your experience, how helpful was (were)..." (very or moderately helpful), n (%)	Qualitative results (n=14); illustrative quotes
Physical activity tracker	24 (73)	• "I wore the activity tracker every day and I thought it was really helpful in monitoring my activity."
Mobile health website	19 (58)	• "I think also the visualization of the [website] was just helpful to get a sense of just how, I guess statistically how my body was changing." • "I also felt like the website, the STRIDE website, it didn't—wasn't—it didn't feel particularly mobile-friendly, and I looked at it on my phone."
Coach calls	26 (79)	• "I thought the lifestyle coach was great, also. [...] She was awesome, very concrete, and actually gave me usable advice versus just preaching."
Digital scale	32 (97)	• "I liked the scale, the wireless scale, and having that linked to my phone so that I had kind of that instant feedback." • "I liked that there was always [something] to keep me accountable, the scale to keep me accountable, you know"
Text messages	24 (73)	• "I enjoyed the congratulatory texts; they just made you feel good. The reminders were nice, especially if it was a little bit more to go, then it would give me that extra spur to take a look or something like that."

Overall, the intervention was well received, and focus group participants reported that the mHealth lifestyle intervention helped them in 2 main ways. First, it promoted accountability. Second, the intervention motivated them and provided tangible support toward their PA goals. Similarly, the most highly rated intervention components (ie, digital scale, coach calls, PA tracker, and text messages) were cited as the most useful mechanisms for both accountability and motivation. The mHealth website was viewed more as a place to see all activities tracked in one place but less as a mechanism toward promoting accountability or motivation toward goals. Participants reported enjoying the mHealth lifestyle intervention program and believed that it led to more PA and less GWG. For example, participants said the following:

> *I think just the act of daily weighing and just paying attention to steps, especially I knew I was going to be going to the gym and working out for pregnancy, but the counting of the steps and the daily weigh-ins really quantified and made me pay attention to what I was doing. And I definitely walked around more and took the stairs and other stuff than I would have otherwise*

> *My first pregnancy, I gained a lot more weight than with this one, and I credit a lot of it due to the STRIDE study, just being mindful and encouraged to be healthier and more active, I gained much less, significantly. So I was really grateful to be a part of it and I'm really glad I got to do it.*

These quotes highlight the value of STRIDE to participants and how the tools provided by the intervention facilitated self-monitoring and improved health behaviors, including PA.

The focus group participants also mentioned potential improvements for the STRIDE intervention. Many participants felt that more nutrition information would have helped them achieve their GWG goals better. Ultimately, the study recommended, but did not require, popular free mHealth apps and websites such as MyFitnessPal and Choose MyPlate. Although there was mixed feedback on the use of these additional tools, many participants wanted more guidance on nutrition that was tailored to their actual dietary habits. For example, participants said the following:

> *I felt like the nutrition piece was nice to have as part of the overall study but wasn't really integral in what we were doing, even though [...] that actually is a big factor in your weight, generally speaking. [...] Like, I think just more support in that area, but I don't know that that support had to be necessarily talking to someone more than once a month. But I felt like it was an afterthought of the study, because it wasn't even built into the program that you would track your eating.*

> *I think where it fell down for me was a little bit around the nutrition, and because we weren't really tracking that as part of the study, it felt like it was nice to have. So that—the orientation for the individual was more around the, like, eating piece. But we weren't really tracking that, like.*

Participants also wanted the mobile website to be a smartphone app. Many participants stated that they primarily used their mobile phones to access websites. Therefore, having an app or a more mobile-friendly website would facilitate website use.

Exploratory Analyses

Although this pilot study was not powered for clinical outcomes, in exploratory analyses, we found that participants in the intervention had greater change in total activity per week compared with that in usual care (Table 3). Participants in the intervention arm had a 23.46 MET hours greater change in self-reported total PA per week (95% CI 1.13 to 45.8) and a 247.2-minute greater change in moderate intensity PA per week (95% CI 36.2 to 530.6) in unadjusted models, but this effect was attenuated in adjusted models (change in total PA: 15.55 MET hours per week, 95% CI −6.32 to 37.42; change in moderate intensity PA: 199.6 minutes per week, 95% CI −43.7 to 442.9). We found no difference between arms in total GWG (mean difference 1.14 kg, 95% CI −0.71 to 3.00) or rate of GWG (mean difference 0.03 kg, 95% CI −0.02 to 0.09). In addition, we did not find, and were not powered to find, any significant differences between intervention and usual care study arms in any adverse perinatal outcomes (data not shown).

Table 3. Change in physical activity and gestational weight gain (GWG) by treatment condition.

	Intervention (N=33), mean (SD)	Usual care (N=35), mean (SD)	Unadjusted model, mean difference (95% CI)	Adjusted model[a], mean difference (95% CI)
Activity in metabolic equivalent of task hours per week				
Total activity			23.46 (1.13 to 45.80)	15.55 (−6.32 to 37.42)
8- to 15-week gestation	105.3 (62.7)	135.8 (100.2)		
33- to 36-week gestation	108.9 (62.0)	115.9 (81.3)		
Change	3.6 (36.8)	−19.9 (53.0)		
Moderate activity			13.56 (−2.19 to 29.30	11.00 (−3.15 to 25.15)
8- to 15-week gestation	49.0 (35.6)	65.7 (62.5)		
33- to 36-week gestation	52.7 (33.8)	55.9 (51.4)		
Change	3.7 (23.4)	−9.8 (39.0)		
Vigorous activity			0.25 (−2.10 to 2.60)	0.25 (−2.10 to 2.60)
8- to 15-week gestation	3.1 (4.8)	4.1 (6.1)		
33- to 36-week gestation	1.0 (1.5)	1.7 (3.7)		
Change	−2.1 (4.7)	−2.4 (4.9)		
Sports and exercise			2.96 (−3.20 to 9.12)	2.46 (−2.23 to 7.16)
8- to 15-week gestation	15.0 (14.8)	17.3 (21.4)		
33- to 36-week gestation	12.6 (11.2)	12.0 (4.3)		
Change	−2.4 (12.6)	−5.3 (12.6)		
Moderate sports and exercise			2.71 (−2.06 to 7.47)	2.30 (−1.68 to 6.28)
8- to 15-week gestation	12.0 (12.2)	13.2 (17.3)		
33- to 36-week gestation	11.7 (11.0)	10.3 (12.0)		
Change	−0.3 (10.1)	−3.0 (9.5)		
Activity in minutes per week				
Moderate activity			247.2 (−36.2 to 530.6)	199.6 (−43.7 to 442.9)
8- to 15-week gestation	818.5 (631.2)	1113.4 (1121.6)		
33- to 36-week gestation	888.3 (596.2)	936.1 (872.9)		
Change	69.9 (393.9)	−177.4 (716.8)		
Sports and exercise			43.67 (−31.5 to 118.9)	35.98 (−23.3 to 95.23)
8- to 15-week gestation	194.2 (173.8)	222.8 (267.9)		
33- to 36-week gestation	183.8 (148.1)	168.7 (179.2)		
Change	−10.4 (152.0)	−54.1 (156.1)		
GWG				
Total GWG (kg[b])	12.7 (3.8)	12.1 (4.1)	0.61 (−1.35 to 2.57)	1.14 (−0.71 to 3.00)
Rate of total GWG (kg/week)	0.4 (0.1)	0.4 (0.1)	0.01 (−0.05 to 0.07)	0.03 (−0.02 to 0.09)

[a]Adjusted model for change in physical activity adjusted for physical activity at baseline survey, age, parity, prepregnancy BMI, race or ethnicity, and difference in gestational age (GA) weeks between baseline survey and the second survey. Adjusted model for GWG adjusted for age, parity, prepregnancy BMI, race or ethnicity, and difference in GA weeks between the last and first measured weight during pregnancy.

[b]Total GWG was defined as the last measured weight within 3 weeks before delivery minus the first measured weight after conception and up to 13 weeks of GA. Rate of total GWG was defined as total GWG divided by difference in GA weeks between the first and last measured weight during pregnancy.

Discussion

Principal Findings

In this mixed methods acceptability and feasibility randomized controlled pilot trial, we found that an mHealth intervention for pregnant patients with overweight or obesity was feasible and acceptable for participants and successfully promoted weight and PA self-monitoring. There was a high level of adherence to self-monitoring of weight and PA among participants in the intervention arm. and overall, the participants rated the program highly. Focus groups found that participants desired more support related to nutrition and a more mobile-friendly app instead of an mHealth website. In exploratory analyses, we found that the mHealth lifestyle intervention increased minutes of PA per week compared with usual care, but there was no difference in GWG. mHealth interventions with more nutrition support are likely needed to effectively affect GWG. It is important to note that although there are modifiable lifestyle factors (eg, nutrition and PA) that contribute to obesity, it is recognized as a complex, chronic disease driven by biological, genetic, environmental, and socioeconomic factors. Therefore, despite engagement with effective lifestyle interventions, pregnancy weight gain may differ among individuals because of a variety of factors outside of the scope of interventions, which may result in null intervention findings.

Comparison With Prior Work

The findings of this study contribute to a small but growing body of literature with mixed results on mHealth interventions to improve PA in pregnant patients with overweight or obesity. Various pilot studies assessing the use of PA trackers (eg, Fitbit) to increase PA among pregnant patients have found no or small overall increases in steps [48-50]. However, Ainscough et al [51] found that an mHealth intervention, delivered via a smartphone app and grounded in behavior change techniques, increased motivation to engage in exercise, self-reported total PA (MET minutes per week), and moderate intensity PA (minutes per week) compared with the control group [51]. This suggests that mHealth interventions to increase PA among pregnant patients may be more effective when using behavior change theories and techniques. Our findings also point to the need for more research to better understand how to maximize the effectiveness of mHealth interventions in this population.

Our study also provides some additional exploratory evidence on the effects of mHealth lifestyle interventions on GWG. Multiple effective components (ie, daily self-monitoring of weight, PA, nutrition, goal setting, feedback, reinforcement, and problem solving) are likely needed to improve outcomes in pregnant patients with overweight or obesity. The lack of nutrition focus and food self-monitoring in our intervention could have contributed to null GWG findings. In a systematic review and meta-analysis of 11 exclusively digital interventions to encourage PA, appropriate weight gain during pregnancy, and healthy eating among pregnant patients, researchers found no overall benefit of exclusively digital interventions on GWG. However, effective individual interventions had twice as many behavior change techniques from *feedback and monitoring* domains and *goals and planning* domains than ineffective interventions. Moreover, higher user engagement with key behavior change techniques had a positive association with greater intervention effectiveness. Overall, effective interventions used both more behavior change techniques and interactivity in the form of personalized feedback, prompts to remind participants to use behavior change techniques and messages of encouragement, similar to our intervention [52]. Another systematic review and meta-analysis of 21 randomized controlled trials assessing the effects of technology-supported interventions on GWG found that these interventions had small effects on GWG, energy intake, eating behaviors, and PA. However, technology-supported interventions that included tracking tools, daily monitoring using devices, and face-to-face sessions were associated with slightly larger effects, particularly for PA [53].

Taken together, our findings demonstrate opportunities to (1) leverage technology to facilitate adherence to self-monitoring via automated, real-time transmission of weight and PA self-monitoring data, including real-time feedback on GWG in relation to the IOM guidelines and (2) incorporate tailored feedback from health care professionals (a lifestyle coach).

Limitations

Our study had several limitations. First, this was a small randomized controlled pilot trial to assess the acceptability and feasibility of the intervention. Therefore, our study was not powered to detect clinical outcomes. Second, we relied on self-reported measures of PA, although PPAQ is a validated self-report tool for assessing PA during pregnancy. Third, pregnant patients randomized to the usual care arm had higher levels of baseline PA than those randomized to the intervention arm; however, we adjusted for baseline differences in our analysis. Fourth, this study used BMI measurements, which on their own, have limitations. However, these were the data available in the EHRs, and BMI is still widely used clinically because of its ease of measurement. Finally, the participants were not masked to the study group, which could have led to biased reports in the intervention arm.

Conclusions

Our study demonstrated that the use of mHealth technology to deliver a theory-based lifestyle intervention is acceptable for pregnant patients with overweight or obesity. One goal of this pilot trial was to inform the design of a larger randomized controlled trial. To this end, the study team is currently implementing a pragmatic cluster randomized clinical trial that expands the pilot intervention based on participant feedback to incorporate additional nutrition support via nutrition self-monitoring and implements an app-based version of the pilot mHealth website to support better use. More effective interventions with broader reach are needed to help pregnant patients with overweight or obesity increase their PA and meet the IOM guidelines for GWG.

Acknowledgments

This study was funded by the Delivery Science Grants Program of the Permanente Medical Group. TT received funding from the Permanente Medical Group Delivery Science Fellowship Program and the National Institute of Diabetes and Digestive and Kidney Diseases (grant T32DK11668401). SEB received funding from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (grant K99HD100585) and Kaiser Permanente Northern California Community Health. MH received funding from R01DK118455 for this study. SDB received funding from the National Institute of Diabetes and Digestive and Kidney Diseases (grant K01DK099404).

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 9707 KB - formative_v6i6e33929_app1.pdf]

References

1. Hedderson MM, Gunderson EP, Ferrara A. Gestational weight gain and risk of gestational diabetes mellitus. Obstet Gynecol 2010 Mar;115(3):597-604 [FREE Full text] [doi: 10.1097/AOG.0b013e3181cfce4f] [Medline: 20177292]
2. Vesco KK, Sharma AJ, Dietz PM, Rizzo JH, Callaghan WM, England L, et al. Newborn size among obese women with weight gain outside the 2009 Institute of Medicine recommendation. Obstet Gynecol 2011 Apr;117(4):812-818. [doi: 10.1097/AOG.0b013e3182113ae4] [Medline: 21422851]
3. Chu SY, Callaghan WM, Bish CL, D'Angelo D. Gestational weight gain by body mass index among US women delivering live births, 2004-2005: fueling future obesity. Am J Obstet Gynecol 2009 Mar;200(3):271.e1-271.e7. [doi: 10.1016/j.ajog.2008.09.879] [Medline: 19136091]
4. Oken E, Rifas-Shiman SL, Field AE, Frazier AL, Gillman MW. Maternal gestational weight gain and offspring weight in adolescence. Obstet Gynecol 2008 Nov;112(5):999-1006 [FREE Full text] [doi: 10.1097/AOG.0b013e31818a5d50] [Medline: 18978098]
5. Rogers EM, Collins-Jarvis L, Schmitz J. The PEN project in Santa Monica: interactive communication, equality, and political action. J Am Soc Inf Sci 1994 Jul;45(6):401-410. [doi: 10.1002/(sici)1097-4571(199407)45:6<401::aid-asi6>3.0.co;2-n]
6. Rogers EM. Communication Technology: The New Media in Society. New York, NY, USA: Free Press; 1986.
7. Schmitz J, Rogers EM, Phillips K, Paschal D. The public electronic network (PEN) and the homeless in Santa Monica. J Appl Commun Res 1995 Feb;23(1):26-43. [doi: 10.1080/00909889509365412]
8. Askov EN, Jo Clark C. Using computers in adult literacy instruction. J Read 1991 Mar;34(6):434-448.
9. Bell JA. The role of microcomputers in patient education. Comput Nurs 1986;4(6):255-258. [Medline: 3539286]
10. Chen LL. Digital multimedia instruction: past, present, and future. J Educ Technol Syst 1994 Jan 01;23(2):169-175. [doi: 10.2190/ylu6-13dh-yf3u-5q9t]
11. Green MJ, Fost N. An interactive computer program for educating and counseling patients about genetic susceptibility to breast cancer. J Cancer Educ 1997;12(4):204-208. [doi: 10.1080/08858199709528490] [Medline: 9440011]
12. Luker KA, Caress AL. Evaluating computer assisted learning for renal patients. Int J Nurs Stud 1992 Aug;29(3):237-250. [doi: 10.1016/0020-7489(92)90027-e] [Medline: 1517025]
13. Barker DI. A technological revolution in higher education. J Educ Technol Syst 1994 Jan 01;23(2):155-168. [doi: 10.2190/5dh4-nupq-j8lu-r6ve]
14. Garrud P, Chapman IR, Gordon SA, Herbert M. Non-verbal communication: evaluation of a computer-assisted learning package. Med Educ 1993 Nov;27(6):474-478. [doi: 10.1111/j.1365-2923.1993.tb00306.x] [Medline: 8208153]
15. Graziadei WD, McCombs GM. The 21st century classroom-scholarship environment: what will it be like? J Educ Technol Syst 1995 Dec 01;24(2):97-112. [doi: 10.2190/kghh-brhb-0m99-lv94]
16. Barakat R, Refoyo I, Coteron J, Franco E. Exercise during pregnancy has a preventative effect on excessive maternal weight gain and gestational diabetes. A randomized controlled trial. Braz J Phys Ther 2019;23(2):148-155 [FREE Full text] [doi: 10.1016/j.bjpt.2018.11.005] [Medline: 30470666]
17. Dipietro L, Evenson KR, Bloodgood B, Sprow K, Troiano RP, Piercy KL, 2018 Physical Activity Guidelines Advisory Committe*. Benefits of physical activity during pregnancy and postpartum: an umbrella review. Med Sci Sports Exerc 2019 Jun;51(6):1292-1302 [FREE Full text] [doi: 10.1249/MSS.0000000000001941] [Medline: 31095086]
18. Ruchat SM, Mottola MF, Skow RJ, Nagpal TS, Meah VL, James M, et al. Effectiveness of exercise interventions in the prevention of excessive gestational weight gain and postpartum weight retention: a systematic review and meta-analysis. Br J Sports Med 2018 Nov;52(21):1347-1356. [doi: 10.1136/bjsports-2018-099399] [Medline: 30337461]

19. Streuling I, Beyerlein A, Rosenfeld E, Hofmann H, Schulz T, von Kries R. Physical activity and gestational weight gain: a meta-analysis of intervention trials. BJOG 2011 Feb;118(3):278-284. [doi: 10.1111/j.1471-0528.2010.02801.x] [Medline: 21134106]
20. The American College of Obstetricians and Gynecologists. Exercise during pregnancy and the postpartum period. ACOG Technical Bulletin Number 189--February 1994. Int J Gynaecol Obstet 1994 Apr;45(1):65-70. [Medline: 7913067]
21. Evenson KR, Wen F. National trends in self-reported physical activity and sedentary behaviors among pregnant women: NHANES 1999-2006. Prev Med 2010 Mar;50(3):123-128. [doi: 10.1016/j.ypmed.2009.12.015] [Medline: 20053370]
22. de Vries HJ, Kooiman TJ, van Ittersum MW, van Brussel M, de Groot M. Do activity monitors increase physical activity in adults with overweight or obesity? A systematic review and meta-analysis. Obesity (Silver Spring) 2016 Oct;24(10):2078-2091 [FREE Full text] [doi: 10.1002/oby.21619] [Medline: 27670401]
23. Ambeba EJ, Ye L, Sereika SM, Styn MA, Acharya SD, Sevick MA, et al. The use of mHealth to deliver tailored messages reduces reported energy and fat intake. J Cardiovasc Nurs 2015;30(1):35-43 [FREE Full text] [doi: 10.1097/JCN.0000000000000120] [Medline: 24434827]
24. Grant L, Evans A. Principles of Behavior Analysis. New York, NY, USA: HarperCollins Publishers; 1994.
25. Patterson R. Changing Patient Behavior: Improving Outcomes in Health and Disease Management. San Francisco, CA, USA: Jossey-Bass; 2001.
26. Burke LE, Conroy MB, Sereika SM, Elci OU, Styn MA, Acharya SD, et al. The effect of electronic self-monitoring on weight loss and dietary intake: a randomized behavioral weight loss trial. Obesity (Silver Spring) 2011 Feb;19(2):338-344 [FREE Full text] [doi: 10.1038/oby.2010.208] [Medline: 20847736]
27. Steinberg DM, Tate DF, Bennett GG, Ennett S, Samuel-Hodge C, Ward DS. The efficacy of a daily self-weighing weight loss intervention using smart scales and e-mail. Obesity (Silver Spring) 2013 Sep;21(9):1789-1797 [FREE Full text] [doi: 10.1002/oby.20396] [Medline: 23512320]
28. Institute of Medicine (US) and National Research Council (US) Committee to Reexamine IOM Pregnancy Weight Guidelines, Rasmussen KM, Yaktine AL. Weight Gain During Pregnancy: Reexamining the Guidelines. Washington, DC, USA: National Academies Press (US); 2009.
29. Diabetes Prevention Program (DPP) Research Group. The Diabetes Prevention Program (DPP): description of lifestyle intervention. Diabetes Care 2002 Dec;25(12):2165-2171 [FREE Full text] [doi: 10.2337/diacare.25.12.2165] [Medline: 12453955]
30. Ferrara A, Hedderson MM, Brown SD, Ehrlich SF, Tsai AL, Feng J, et al. A telehealth lifestyle intervention to reduce excess gestational weight gain in pregnant women with overweight or obesity (GLOW): a randomised, parallel-group, controlled trial. Lancet Diabetes Endocrinol 2020 Jun;8(6):490-500 [FREE Full text] [doi: 10.1016/S2213-8587(20)30107-8] [Medline: 32445736]
31. Bandura A. Social Learning Theory. Englewood Cliffs, NJ, USA: Prentice Hall; 1977.
32. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. Hoboken, NJ, USA: Prentice-Hall; 1986.
33. Bandura A. Self-Efficacy: The Exercise of Control. New York, NY, USA: W H Freeman/Times Books/Henry Holt; 1997.
34. Prochaska JO, DiClemente CC. Common processes of self-change in smoking, weight control, and psychological distress. In: Sniff-man S, Wills TA, editors. Coping and Substance Use. San Diego, CA, USA: Academic Press; 1985:345-363.
35. Brown SD, Hedderson MM, Ehrlich SF, Galarce MN, Tsai AL, Quesenberry CP, et al. Gestational weight gain and optimal wellness (GLOW): rationale and methods for a randomized controlled trial of a lifestyle intervention among pregnant women with overweight or obesity. BMC Pregnancy Childbirth 2019 Apr 30;19(1):145 [FREE Full text] [doi: 10.1186/s12884-019-2293-8] [Medline: 31039753]
36. Ferrara A, Hedderson MM, Albright CL, Ehrlich SF, Quesenberry Jr CP, Peng T, et al. A pregnancy and postpartum lifestyle intervention in women with gestational diabetes mellitus reduces diabetes risk factors: a feasibility randomized control trial. Diabetes Care 2011 Jul;34(7):1519-1525 [FREE Full text] [doi: 10.2337/dc10-2221] [Medline: 21540430]
37. Ferrara A, Hedderson MM, Brown SD, Albright CL, Ehrlich SF, Tsai AL, et al. The comparative effectiveness of diabetes prevention strategies to reduce postpartum weight retention in women with gestational diabetes mellitus: the gestational diabetes' effects on moms (GEM) cluster randomized controlled trial. Diabetes Care 2016 Jan;39(1):65-74 [FREE Full text] [doi: 10.2337/dc15-1254] [Medline: 26657945]
38. Linde JA, Jeffery RW, French SA, Pronk NP, Boyle RG. Self-weighing in weight gain prevention and weight loss trials. Ann Behav Med 2005 Dec;30(3):210-216. [doi: 10.1207/s15324796abm3003_5] [Medline: 16336072]
39. Wing RR, Tate DF, Gorin AA, Raynor HA, Fava JL. A self-regulation program for maintenance of weight loss. N Engl J Med 2006 Oct 12;355(15):1563-1571. [doi: 10.1056/NEJMoa061883] [Medline: 17035649]
40. Ferguson T, Rowlands AV, Olds T, Maher C. The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. Int J Behav Nutr Phys Act 2015 Mar 27;12:42 [FREE Full text] [doi: 10.1186/s12966-015-0201-9] [Medline: 25890168]
41. O'Connell S, ÓLaighin G, Kelly L, Murphy E, Beirne S, Burke N, et al. These shoes are made for walking: sensitivity performance evaluation of commercial activity monitors under the expected conditions and circumstances required to

achieve the international daily step goal of 10,000 steps. PLoS One 2016 May 11;11(5):e0154956 [FREE Full text] [doi: 10.1371/journal.pone.0154956] [Medline: 27167121]
42. O'Connell S, ÓLaighin G, Quinlan LR. When a step is not a step! Specificity analysis of five physical activity monitors. PLoS One 2017 Jan 13;12(1):e0169616 [FREE Full text] [doi: 10.1371/journal.pone.0169616] [Medline: 28085918]
43. Zhao Y, Parvinzamir F, Wei H, Liu E, Deng Z, Dong F, et al. Visual analytics for health monitoring and risk management in CARRE. In: Proceedings of the 10th International Conference on E-Learning and Games. 2016 Presented at: Edutainment '16; April 14-16, 2016; Hangzhou, China p. 380-391. [doi: 10.1007/978-3-319-40259-8]
44. Chasan-Taber L, Schmidt MD, Roberts DE, Hosmer D, Markenson G, Freedson PS. Development and validation of a Pregnancy Physical Activity Questionnaire. Med Sci Sports Exerc 2004 Oct;36(10):1750-1760. [doi: 10.1249/01.mss.0000142303.49306.0d] [Medline: 15595297]
45. Roberts DE, Fragala MS, Pober D, Chasan-Taber L, Freedson PS. Energy cost of physical activities during pregnancy. Med Sci Sports Exerc 2002;34(5):S124. [doi: 10.1097/00005768-200205001-00701]
46. Ainsworth BE, Haskell WL, Herrmann SD, Meckes N, Bassett Jr DR, Tudor-Locke C, et al. 2011 Compendium of Physical Activities: a second update of codes and MET values. Med Sci Sports Exerc 2011 Aug;43(8):1575-1581. [doi: 10.1249/MSS.0b013e31821ece12] [Medline: 21681120]
47. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
48. Choi J, Lee JH, Vittinghoff E, Fukuoka Y. mHealth physical activity intervention: a randomized pilot study in physically inactive pregnant women. Matern Child Health J 2016 May;20(5):1091-1101 [FREE Full text] [doi: 10.1007/s10995-015-1895-7] [Medline: 26649879]
49. Darvall JN, Wang A, Nazeem MN, Harrison CL, Clarke L, Mendoza C, et al. A pedometer-guided physical activity intervention for obese pregnant women (the Fit MUM Study): randomized feasibility study. JMIR Mhealth Uhealth 2020 May 26;8(5):e15112 [FREE Full text] [doi: 10.2196/15112] [Medline: 32348280]
50. Larsen B, Micucci S, Hartman SJ, Ramos G. Feasibility and acceptability of a counseling- and mHealth-based physical activity intervention for pregnant women with diabetes: the fit for two pilot study. JMIR Mhealth Uhealth 2020 Oct 21;8(10):e18915 [FREE Full text] [doi: 10.2196/18915] [Medline: 33084584]
51. Ainscough KM, O'Brien EC, Lindsay KL, Kennelly MA, O'Sullivan EJ, O'Brien OA, et al. Nutrition, behavior change and physical activity outcomes from the PEARS RCT-an mHealth-supported, lifestyle intervention among pregnant women with overweight and obesity. Front Endocrinol (Lausanne) 2020 Feb 4;10:938 [FREE Full text] [doi: 10.3389/fendo.2019.00938] [Medline: 32117047]
52. Rhodes A, Smith AD, Chadwick P, Croker H, Llewellyn CH. Exclusively digital health interventions targeting diet, physical activity, and weight gain in pregnant women: systematic review and meta-analysis. JMIR Mhealth Uhealth 2020 Jul 10;8(7):e18255 [FREE Full text] [doi: 10.2196/18255] [Medline: 32673251]
53. Leonard KS, Evans MB, Oravecz Z, Smyth JM, Symons Downs D. Effect of technology-supported interventions on prenatal gestational weight gain, physical activity, and healthy eating behaviors: a systematic review and meta-analysis. J Technol Behav Sci 2021;6(1):25-41. [doi: 10.1007/s41347-020-00155-6]

Abbreviations

EHR: electronic health record
GA: gestational age
GLOW: Gestational Weight Gain and Optimal Wellness
GWG: gestational weight gain
IOM: Institute of Medicine
KPNC: Kaiser Permanente Northern California
MET: metabolic equivalent of task
mHealth: mobile health
MVPA: moderate to vigorous intensity physical activity
PA: physical activity
PPAQ: Pregnancy Physical Activity Questionnaire
STRIDE: Study of a Randomized Intervention Designed to Increase Exercise in Pregnancy

Edited by T Leung, A Mavragani; submitted 29.09.21; peer-reviewed by B Larsen, M Graham, L Tully; comments to author 04.01.22; revised version received 29.03.22; accepted 22.04.22; published 22.06.22.

Please cite as:
Thomas T, Xu F, Sridhar S, Sedgwick T, Nkemere L, Badon SE, Quesenberry C, Ferrara A, Mandel S, Brown SD, Hedderson M
A Web-Based mHealth Intervention With Telephone Support to Increase Physical Activity Among Pregnant Patients With Overweight or Obesity: Feasibility Randomized Controlled Trial
JMIR Form Res 2022;6(6):e33929
URL: https://formative.jmir.org/2022/6/e33929
doi:10.2196/33929
PMID:35731565

Original Paper

Predictors of Disengagement and Symptom Improvement Among Adults With Depression Enrolled in Talkspace, a Technology-Mediated Psychotherapy Platform: Naturalistic Observational Study

Doyanne Darnell[1], PhD; Michael D Pullmann[1], PhD; Thomas D Hull[2], PhD; Shiyu Chen[1], MPH; Patricia Areán[1], PhD

[1]Department of Psychiatry & Behavioral Sciences, University of Washington, Seattle, WA, United States
[2]Talkspace, Seattle, WA, United States

Corresponding Author:
Doyanne Darnell, PhD
Department of Psychiatry & Behavioral Sciences
University of Washington
325 Ninth Avenue
Box 359911
Seattle, WA, 98104
United States
Phone: 1 206 744 9108
Email: darnelld@uw.edu

Abstract

Background: Depression is a common psychiatric condition with an estimated lifetime prevalence for major depression of 16.6% in the US adult population and is effectively treated through psychotherapy. The widespread availability of the internet and personal devices such as smartphones are changing the landscape of delivery of psychotherapy; however, little is known about whether and for whom this type of therapy is beneficial, and whether having synchronous video-based sessions provides additional benefits to clients above and beyond messaging-based therapy.

Objective: This study examined the outcomes associated with the use of a digital platform (Talkspace) for technology-mediated psychotherapy. We examined the duration of client engagement in therapy and client depression score trajectories over 16 weeks. We explored the association of client characteristics, therapist characteristics, and service plan type with time-to-disengagement and trajectories of change in depression scores.

Methods: This naturalistic observational study assessed data collected routinely by the platform between January 2016 and January 2018 and examined psychotherapy outcomes among a large representative sample of adult clients with clinically significant depression. Treatment disengagement was defined as a lack of client-initiated communication for more than 4 weeks. Clients completed the Patient Health Questionnaire-8 item (PHQ-8) at intake and every 3 weeks via an in-app survey. Cox regression analysis was used to examine the time until and predictors of disengagement. Changes in depression scores and predictors of change over time were examined using mixed-effects regression.

Results: The study included 5890 clients and 1271 therapists. Client scores on the PHQ-8 declined over time, with the average client improving from a score of 15 to below the clinical cutoff of 10 by week 6. At the same time point, 37% of clients had disengaged from the therapy. When combined into a final Cox regression model, those who were more likely to disengage were clients aged 18 to 25 years versus those aged ≥50 years (odds ratio [OR] 0.82, 95% CI 0.74-0.9; P<.001), had higher education (OR 1.14, 95% CI 1.06-1.22; P<.001), had been in therapy before (OR 1.09, 95% CI 1.02-1.17; P=.01), and were living with a partner but unmarried versus single (OR 1.14, 95% CI 1.02-1.27; P=.02). Having a therapist with >10 years of experience was related to lower odds of disengagement (OR 0.87, 95% CI 0.8-0.94; P=.01). When combined into a final regression model predicting improvement in depression scores over time, clients showing more improvement were those with an associate's degree or higher (linear estimate=−0.07, P=.002) and higher intake PHQ-8 scores (estimate=3.73, P<.001). There were no differences based on the plan type.

Conclusions: Our findings add to the growing literature showing the benefits of technology-mediated psychotherapy over a relatively brief period (16 weeks).

(JMIR Form Res 2022;6(6):e36521) doi:10.2196/36521

KEYWORDS

depression; psychotherapy; disengagement; internet; web-based; technology-mediated psychotherapy

Introduction

Background

Depression is a common psychiatric condition with an estimated lifetime prevalence of major depression of 16.6% and a previous year prevalence of a major depressive episode of 7.8% to 8.6% among the US adult population [1,2]. Psychotherapy is an effective treatment for depression [3] and is often preferred to medication [4]; however, clients face numerous challenges in obtaining what has traditionally been face-to-face in-person treatment, such as the need to travel to a psychotherapist's office and potential stigma associated with accessing this care [5]. The widespread availability of the internet and personal devices such as smartphones is changing the landscape of delivery of psychotherapy, allowing consumers to access care more flexibly and without the need to go to a clinic [6]. These opportunities continue to grow in the wake of the COVID-19 pandemic, and the resultant increase in web-based health care is initiated in response to social distancing policies [7].

Numerous technology platforms are now available that allow clients to flexibly communicate with a psychotherapist via text messages, voice, or videoconferencing software; this is known as technology-mediated psychotherapy [6]. A common distinction made regarding technology-mediated psychotherapy is whether it is conducted in an asynchronous or synchronous fashion. Asynchronous refers to communication sent and received by therapists and clients at different times, such as a therapist sending a client an email with a homework assignment and then sending feedback on the assignment once completed. Synchronous refers to communication sent and received by therapists and clients at the same time, such as telephone conversations or a web-based chat session.

Prior Work

Over the past 20 years, numerous clinical trials of technology-mediated synchronous and asynchronous psychotherapy have been conducted, most of which have evaluated structured cognitive-behavioral approaches. Findings from meta-analyses based on trials conducted in English of psychotherapy for depression by videoconference or telephone (synchronous) indicate that these forms of therapy are similarly effective to face-to-face in-person treatment and with similar (~80%), if not better, completion rates [8-10]. These meta-analyses reflect diverse samples recruited from a variety of settings (eg, veterans, primary care, HIV clinics, outpatient psychotherapy clinics, and community samples).

Fewer studies on the effectiveness of text- or chat-based therapy (whether synchronous or asynchronous) have been conducted, and the findings have been mixed. One clinical trial evaluating asynchronous cognitive behavior therapy (CBT)–oriented psychotherapy for depression among adults in Sweden assessed the benefit of an 8-week email-based therapy, indicating that email therapy was associated with pre-post benefit but no difference in treatment outcomes as compared with a guided 8-week self-help web-based program or waitlist-control at 6 months [11]. The rates of client completion of the intervention were similarly high between the groups (96% completed the email-based therapy and 93% completed the self-help). Another trial using a web-based platform for chat or text-based psychotherapy found that adult primary care patients in the United Kingdom assigned to a web-based CBT-oriented therapy for 16 weeks experienced greater improvement in depression symptoms than a waitlist or usual primary care comparison group over 8 months [12]. Analyses were based on an intent-to-treat sample, of which a large proportion (48%) did not complete the therapy as intended, according to the therapist's report.

Few studies on technology-mediated psychotherapy have examined how client demographics moderate outcomes or therapy completion. A secondary exploratory analysis of the United Kingdom chat or text-based CBT trial [12] examined moderators of treatment outcome at the 4-month follow-up and found that depression outcomes were better for patients with higher baseline depression severity [13]. In addition, being separated, divorced, or widowed as opposed to being single or married was associated with greater treatment outcomes for those in the CBT condition; age and education level were not significant moderators. Findings from a meta-analysis of a broader array of technology-mediated psychological interventions targeting various mental health diagnoses suggest that client age may be a relevant moderator of outcome; in particular, adults aged 19 to 39 years may experience greater benefits than older or younger clients [14].

Little is known about the effects of technology-mediated psychotherapy delivered routinely, for whom this approach may be beneficial, or whether there is differential effectiveness based on asynchronous versus synchronous delivery. In addition, to date, no studies have compared asynchronous and synchronous technology-mediated psychotherapy. This study aims to fill these gaps by studying outcomes associated with the use of a digital platform (Talkspace) that facilitates technology-mediated psychotherapy either asynchronously or synchronously among adult clients experiencing depression symptoms. Specifically, the platform is a desktop computer and mobile smartphone app that allows 24/7 communication via a secure interactive text chat and voice or video messages with a licensed psychotherapist for a fee [15]. Clients may pay for a plan with optional synchronous 1-hour sessions conducted through chat, telephone, or videoconferencing that can be either 1 or 4 times per month. Several observational studies with selected samples of Talkspace

users suggest that the use of the asynchronous text-based psychotherapy plan is associated with improvement in depression symptoms over 12 to 16 weeks of treatment; however, none of these studies examined predictors of these outcomes and for whom the therapy may be most beneficial [15-18].

Objective

This naturalistic observational study harnesses data collected routinely by the platform and examines psychotherapy outcomes among a large representative sample of clients with clinically significant depression. We examined how long clients remained in therapy before disengaging from the application after their initial enrollment as well as client trajectories of depression scores over 16 weeks after enrollment. In addition, we explored the association between client characteristics, therapist characteristics, and the type of service plan with time-to-disengagement and trajectories of change in depression scores.

Methods

Design and Procedure

This study included longitudinal and observational research with clients who signed up for Talkspace between January 2016 and January 2018 and their therapists using data collected routinely as part of the service. Clients accessed the service through an internet search and either paid out-of-pocket or submitted costs to a private insurance plan for out-of-network reimbursement. The signup process includes a brief intake with a consultation therapist who helps the client select their desired plan (eg, messaging only, monthly video, or weekly video plans) and records the client's therapist preferences, presenting complaints, and demographics. This information is used by the service to identify and offer the client a choice among 3 therapists that most closely match the client's preferences, are licensed in the client's state of residence, and have a successful history of treating conditions the client is reporting. After therapist selection, the client is offered a baseline symptom assessment and can then message their therapist as often as they like. Therapists respond to client messages within 24 hours, gather information for a diagnosis, explain the frame of the medium, and conduct informed consent procedures, after which therapy proceeds 5 days a week.

Therapists encounter the platform through internet searches, professional organizations, and peer contacts. All prospective therapists completed an application process to verify the state licensure, training and degree type, professional insurance, background checks, and confirmation of meeting the National Certification of Quality Assurance standards. Following verification, therapists completed an orientation to the platform, received Health Insurance Portability and Accountability Act, privacy, and security training on the proper use of technological media for delivering care, and took on a small number of training cases with supervision for 30 days before beginning their practice on the platform. Engagement and outcome metrics were monitored after the training phase for quality assurance along with a random peer-review process by other therapists on the platform. The service does not prescribe a specific approach to therapy, and while most therapists (61%) report practicing from a cognitive-behavioral orientation, therapists from many traditions are represented on the platform.

Participants

The participants in this study were clients of the Talkspace service. Inclusion criteria was age 18 years and older, with clinically significant depression as indicated by having a primary diagnosis of depression or a depression-related disorder, and a score of 10 or greater on the 8-item version of the Patient Health Questionnaire (PHQ-8) administered at intake for the service. In addition, given the requirements of the service, participants were English language literate, had internet access, and were able to use mobile or desktop applications for the service. Exclusion criteria for the use of the service were indications during the intake with the consultation therapist, or at any point when working with the treating therapist, of any schizophrenia spectrum and psychotic disorder, or any diagnosis with psychotic features, or any condition requiring hospitalization, or suicidal thoughts or behavior sufficient to be marked a *Yes* on any of questions 3 through 6 on the Columbia Suicide Severity Rating Scale Lifetime-Recent Screen [19] that would require a more intensive level of care than can be offered through an outpatient service.

Measures and Variables

Client Characteristics and Demographics

As part of signing up for the Talkspace service, clients provide their age range (19-25, 26-35, 36-49, 50+ years), gender (female, male, gender queer, transgender male, transgender female, gender variant, other), education level (high school, bachelor's degree or higher), whether they had ever been in therapy before, and their marital status (single, married, living with a partner, divorced, separated, and widowed).

Depression-Related Diagnosis

Clients were assigned a psychiatric diagnosis according to the 5th edition of the Diagnostic and Statistical Manual of Mental Disorders by their matched therapist, which may be assessed through a mix of diagnostic interviewing, often by video, and delivery of standardized symptom assessments. Clients were included in the study if they had a primary diagnosis of a depression-related disorder (eg, major depressive disorder, adjustment disorder with depressed mood, dysthymic disorder, or other mood disorders).

Therapist Characteristics

Therapists provide information about demographics and professional experiences when they apply to join a service network. Therapists can select female, male, gender queer, transgender male, transgender female, gender variant, or other gender. Therapists indicate years of postlicensure experience as psychotherapists (coded as <5 years, 5-10 years, and >10 years) and areas of expertise (for this study, dichotomously coded as whether therapists indicated expertise in treating depression).

Psychotherapy Disengagement

Disengagement from psychotherapy was defined for this study by a lack of client-initiated communication via any means using the Talkspace application (eg, text, voice messages, and videoconferencing sessions) for more than 4 weeks after initial enrollment. These data are drawn from administrative records on exchanges with therapists and are passively collected by the application.

Depression Symptoms

Clients were asked to complete the PHQ-8 [20] at intake and every 3 weeks via an in-app survey. There was an allowance for completing these with a 1-week buffer before or after the deadline. The PHQ-8 includes 8 items that assess the severity of depression symptoms according to the Diagnostic Statistical Manual, Fourth Edition, criteria, with the exception of suicidality or preoccupation with death. Items were rated on a scale from 0 (not at all) to 3 (nearly every day) and summed to obtain a total score. Higher scores indicate greater severity, and a cutoff score of 10 or higher indicates clinically significant depression. A 5-point difference in scores indicates a clinically meaningful change in depression symptoms [21].

Plan Type

Platform clients could opt to purchase one of three service plans: (1) unlimited text, voice, or video messages; (2) unlimited text, voice, or video messages plus once per month 1-hour videoconferencing session; or (3) unlimited text, voice, or video messages plus 4 times per month 1-hour videoconferencing sessions. How often and to what extent clients and therapists communicate varies and is dependent upon the frequency with which clients send messages to their therapist. Therapists are expected to respond to client communications within 24 hours and generally respond within 12 hours, except for therapist days off or other times mutually agreed upon by the therapist and client.

Plan of Analysis

Using R, we examined descriptive statistics for client and therapist characteristics, depression symptoms at intake or baseline, and service plan type that clients were enrolled in. We examined the missingness of study variables at each point of time as described in the description of our results. For the prediction analysis, we collapsed several categorical variables when the sample size within some of the variable categories was low, which made theoretical sense to do so. We collapsed education into high school education versus associate's degree or higher. As only 30 participants were widowed, we collapsed this group into the missing category. We conducted 2 sets of mixed effects regression analyses to examine, (1) trajectories of disengagement over 52 weeks and predictors of disengagement and (2) trajectories of depression scores over 16 weeks and predictors of change in depression over time.

Predicting Psychotherapy Disengagement Over 52 Weeks After Intake

We examined how long clients were in therapy before disengaging over the course of 52 weeks, following their intake with the platform. Furthermore, a duration of 52 weeks was selected based on visual inspection of the data, showing that most participants had stopped communicating with their therapist for more than 4 weeks by that time. Using R, we conducted mixed-effects Cox regression analyses predicting time until disengagement, with random terms for clinician and client to account for nesting of time within client and client within therapist. Model building followed appropriate procedures for our combined confirmatory and exploratory analyses; predictors were first tested individually to conduct a priori confirmatory hypothesis testing, and those that were significant at $P<.05$ were included in a final, combined exploratory model to examine the overall model structure. Significance was tested using the log-rank test for individual variable models. For categorical variables, simple contrast comparisons were used in which each category was compared against the reference category, as described in the results. To estimate the goodness-of-fit, we computed the proportion of the variance accounted for using Nagelkerke R^2 for each model. We also computed the concordance statistic, which provides the fraction of concordant pairs between the model-predicted and actual data and is equal to the receiver operating characteristic curve. A concordance statistic of 0.50 means that the model is no better than random chance, while a statistic of 1.0 would mean there is a perfect concordance between predicted and actual data. Analyses of the relationship between missing data status on demographic and descriptive variables and time until engagement were conducted separately using Cox regression analyses; cases with missing data were excluded from the primary analyses.

Predicting Depression Symptoms Over 16 Weeks After Intake

We examined depression symptoms based on PHQ-8 scores over the course of 16 weeks following the client's service intake. In addition, a duration of 16 weeks was selected based on visual inspection of the data showing that most of the change in scores occurred during the first 16 weeks and missing data were substantial at that point, which is likely primarily owing to high rates of client disengagement from therapy by 16 weeks. Using R, mixed-effects regressions with random terms for therapist and client were computed to adjust standard errors owing to nesting of time within client and client within therapist. The covariance matrix was specified to be unstructured. As described below, the rates of missing data were high for the dependent variable owing to ending treatment or incomplete measures. Missing data were determined to be not missing at random, based on the empirical analyses described below and theoretical reasons. It is highly likely that those who did not complete measures were due to unmeasured reasons, such as treatment attitudes, motivation, and emotional distress or wellness. Therefore, full information maximum likelihood estimation was used for model testing and parameter estimation (to maximize all available data), and variables associated with missing outcome data were included in the final models to statistically adjust for missingness to the greatest extent possible. Findings may only be inferred to the population of clients who completed measures while receiving treatment. Model building followed standard procedures [22]. Model fit was tested by computing the Akaike information criterion, Bayesian information criterion,

and 2-Log Likelihood deviance statistics. The models tested linear, quadratic, and cubic time trends for changes in the PHQ-8 scores over time. The best-fitting models included the linear and quadratic time trends. Then, each potential predictor, including client and therapist characteristics, was entered into individual models, and interaction terms for linear and quadratic time were tested using model fit deviance statistics, with each more complex model tested against simpler models, and all available data were used for each computation. All variables and interaction terms from models that were statistically significant at $P<.05$ were then entered into a final omnibus model. We ran two types of R^2 for each model: the conditional R^2 provides the total proportion of variance accounted for, including fixed and random terms, while the marginal R^2 provides the proportion of variance accounted for by fixed terms only.

Ethics Approval

As part of the terms of service agreement, clients agreed that their anonymized data may be used for research purposes. Considering the data were completely anonymized, this study was exempted from institutional review board approval by the institutional review board of the University of Washington.

Results

Sample Characteristics

The study included 5890 clients and 1271 therapists (Table 1). The participants were primarily female (4504/5890, 76.5%), 26 to 35 years of age (3061/5890, 52%), had a bachelor's degree (3706/5890, 62.9%), were single (3122/5890, 52.8%), and had been in therapy before (4004/5890, 68%). The average PHQ-8 score was 15.2 (SD 3.9). The therapists were primarily female (1114/1271, 87.6%), with more than 5 years of experience (1017/1271, 80%), and close to half endorsed having depression-specific expertise (591/1271, 46.5%). Clients predominantly signed up for a messaging-only Talkspace plan (5389/1271, 91.5%). There were no missing data for client disengagement from therapy; however, we observed modest amounts of missing data for other client and therapist variables (ranging from 0.3% to 13.6%, see Table 1). Missing data analyses found that missing data on the following variables were not significantly associated with the length of time until disengagement: client's gender, client's first time in therapy, or therapist's years of experience. Client age and education were significantly associated, such that those who had missing data on these variables were more likely to disengage (age odds ratio [OR] 1.46, 95% CI 1.15-1.89, $P=.002$; education OR 1.10, 95% CI 1.02-1.19, $P=.013$).

Table 1. Client and therapist characteristics.

Variable	Value
Client: total, n (%)	5890 (100)
Client age (years), n (%)	
18-25	1328 (22.5)
26-35	3061 (52)
36-49	1182 (20.1)
>50	250 (4.2)
Missing	69 (1.2)
Client gender, n (%)	
Female	4505 (76.5)
Other	5 (0.1)
Queer	9 (0.2)
Variant	5 (0.1)
Male	1317 (22.4)
Transgender female	3 (0.1)
Transgender male	7 (0.1)
Missing	39 (0.7)
Client education, n (%)	
High school	1327 (22.5)
Some college	14 (0.2)
Associate's degree	6 (0.1)
Bachelor's degree or higher	3706 (62.9)
Master's degree	21 (0.4)
Doctoral degree	5 (0.1)
Professional degree	8 (0.1)
Missing	803 (13.6)
Client marital status, n (%)	
Divorced	328 (5.6)
Living with a partner	497 (8.4)
Married	1663 (28.2)
Separated	142 (2.4)
Single	3112 (52.8)
Widowed	30 (0.5)
Missing	118 (2.0)
Client first time in therapy	
No, n (%)	4004 (68)
Yes, n (%)	1346 (22.9)
Missing, n (%)	540 (9.2)
Baseline PHQ-8[a], mean (SD)	15.2 (3.9)
Talkspace plan type	
Text only, n (%)	5389 (91.5)
1-month video, n (%)	411 (7)
4-month video, n (%)	90 (1.5)

Variable	Value
Therapist, total (N)	1271 (100)
Therapist gender, n (%)	
Male	157 (12.4)
Female	1114 (87.6)
Therapist years of experience, n (%)	
<5	250 (19.7)
5-10	551 (43.4)
Missing	4 (0.3)
Therapist with depression expertise, n (%)	591 (46.5)

[a]PHQ-8: Patient Health Questionnaire-8 item.

Predicting Psychotherapy Disengagement Over 52 Weeks Post Intake

By week 6, 37% of the clients had disengaged from therapy. Half of the sample disengaged from therapy by week 9 and by week 52, nearly all clients had disengaged (n=5441, 92.4%; Figure 1). Table 2 displays the single-variable predictor models for all the variables that were significantly associated with disengagement. When combined into a final Cox regression model, significant variables included client age, education, whether the client had been in therapy before, marital status, and years of experience as a therapist (Table 3). Compared with those aged 18 to 25 years, those aged 36 to 49 years had 18.9% lower odds of disengaging at any point in time (OR 0.82, 95% CI 0.74-0.9; *P*<.001; Figure 2) and those aged >50 years had 30.1% lower odds of disengaging (OR 0.70, 95% CI 0.59-0.83; *P*<.001). Those with higher education had 13.5% greater odds of disengaging than those with lower education (OR 1.14, 95% CI 1.06-1.22; *P*<.001), and those who were in therapy for the first time had 8.9% greater odds of disengaging (OR 1.09, 95% CI 1.02-1.17; *P*=.01). Participants who were living with a partner had 13.9% greater odds of disengaging as compared with those who were single (OR 1.14, 95% CI 1.02-1.27; *P*=.02); there were no significant differences between those who were single and those who were married, separated, or divorced. Finally, clients with therapists who had more than 10 years of experience had 13.2% lower odds of disengaging (OR 0.87, 95% CI 0.8-0.94; *P*=.01). Although significant in the single-variable predictor model, there were no significant differences in the likelihood of disengagement for clients based on the therapist endorsement of depression-specific expertise in the combined Cox regression model. Plan type, intake PHQ-8 score, client gender, and therapist gender were not included in the final model, as these variables were not significant in the single-variable predictor models. Model fit statistics revealed a very low proportion of variance accounted for and concordance across all models (Tables 3 and 4). The proportion of variance accounted for in the final model was only 1.8%, and the concordance statistic of 0.54 indicated that the model predicted the length of time until disengagement only slightly better than chance.

Figure 1. Time until disengagement.

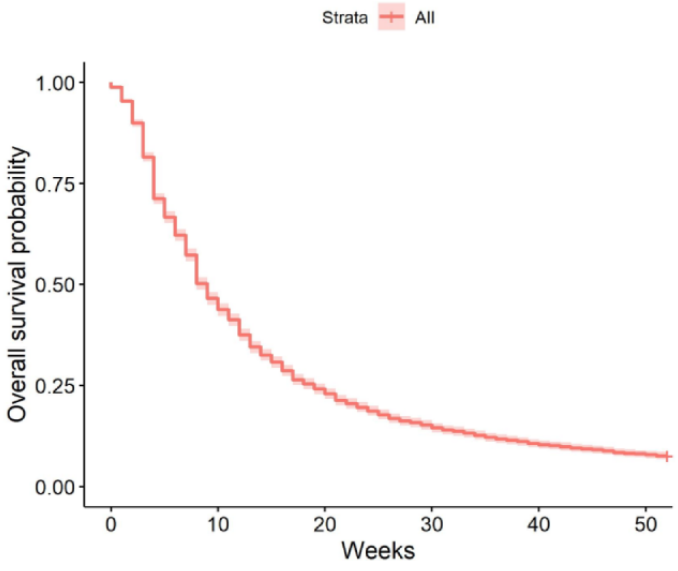


Table 2. Significant single-variable predictor Cox regression models testing time until treatment disengagement[a].

	Median (weeks)	OR[b] (95% CI)	*P* value	Log rank	*P* value
Age (years)				33.78	<.001
18-25	8	Ref	Ref		
26-35	9	0.93 (0.87-0.99)	.035		
36-49	10	0.83 (0.77-0.91)	<.001		
>50	11	0.71 (0.61-0.82)	<.001		
Education				17.24	<.001
High school	9	Ref	Ref		
Associate's degree or higher	8	1.15 (1.08-1.22)	<.001		
First time in treatment				11.37	<.001
No	9	Ref	Ref		
Yes	8	1.12 (1.05-1.19)	<.001		
Marital status				15.31	.004
Single	9	Ref	Ref		
Living with partner	8	1.18 (1.07-1.3)	.001		
Married	9	0.96 (0.91-1.03)	.22		
Separated	10	0.98 (0.82-1.17)	.83		
Divorced	9	0.95 (0.84-1.06)	.37		
Therapist experience				16.45	<.001
<5 years	8	Ref	Ref		
5-10 years	9	0.93 (0.86-1)	.07		
>10 years	9	0.86 (0.8-0.93)	.002		
Therapist with depression-specific expertise				7.97	.004
Not endorsed	8	Ref	Ref		
Endorsed	9	0.92 (0.88-0.98)	.03		

[a]Client n=5890, therapist n=1271. Variables that were tested and were not significant included plan type, intake PHQ-8 score, client gender, and therapist gender. Widowed data were collapsed with missing marital status data. Disengagement from treatment was defined as >4 weeks without any client-initiated communication to the therapist, such as texting, video exchanges, images, or audio clips. The modeling included up to 52 weeks of therapy.

[b]OR: odds ratio.

Table 3. Model fit statistics for individual and final combined nested Cox regression model predicting time until treatment disengagement[a].

	Nagelkerke R^2	Concordance
Individual models		
Age	0.0065	0.52
Education	0.0036	0.52
First time in therapy	0.0022	0.51
Marital status	0.0028	0.51
Therapist experience	0.0030	0.51
Therapist with depression-specific expertise	0.0015	0.51
Final model	0.018	0.54

[a]Disengagement from treatment was defined as >4 weeks without any client-initiated communication to the therapist, such as texting, video exchanges, images, or audio clips. The modeling included up to 52 weeks of therapy.

Figure 2. Time until disengagement stratified by age group.

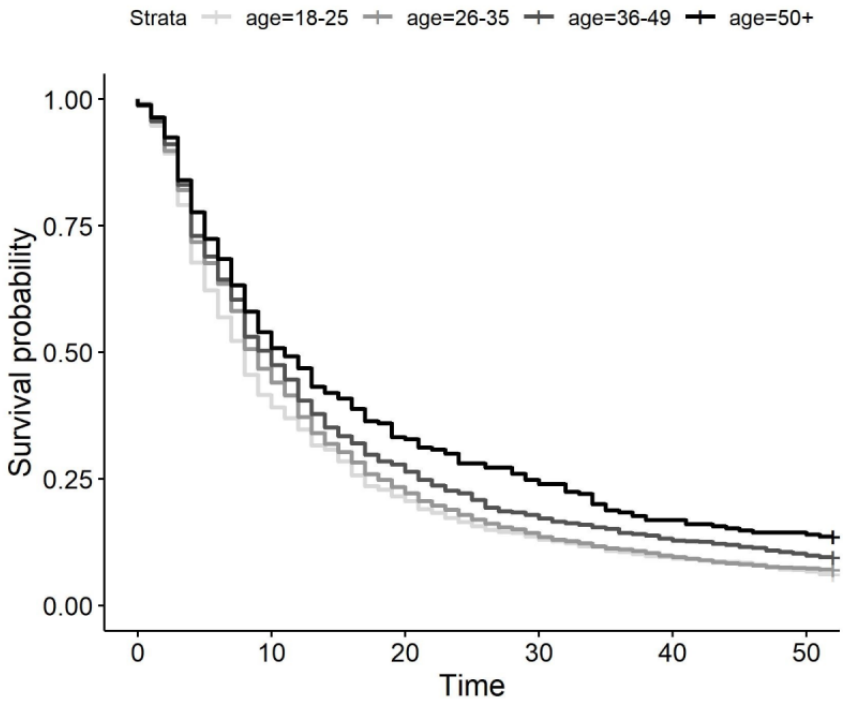


Table 4. Final combined nested Cox regression model predicting time until treatment disengagement[a].

	Coefficient	OR[b] (95% CI)	*P* value
Age (years; ref=18-25)			
26-35	−0.05	0.96 (0.88-1.03)	.27
36-49	−0.20	0.82 (0.74-0.9)	<.001
>50	−0.36	0.7 (0.59-0.83)	<.001
Education (ref=high school)			
Associate's or bachelor's degree or higher	0.13	1.14 (1.06-1.22)	<.001
First time in treatment	0.09	1.09 (1.02-1.17)	.011
Marital status (ref=single)			
Living with a partner	0.13	1.14 (1.02-1.27)	.019
Married	0.03	1.03 (0.95-1.1)	.48
Separated	0.006	1.01 (0.82-1.23)	.95
Divorced	0.04	1.04 (0.91-1.2)	.55
Therapist experience (years; ref=<5)			
5-10	−0.09	0.92 (0.84-0.99)	.06
>10	−0.14	0.87 (0.8-0.94)	.01
Therapist with depression-specific expertise	−0.06	0.94 (0.89-1)	.14

[a]Disengagement from treatment was defined as >4 weeks without any client-initiated communication to the therapist, such as texting, video exchanges, images, or audio clips. The modeling included up to 52 weeks of therapy. Predictor variables that were significant in simpler models were also included.
[b]OR: odds ratio.

Predicting Depression Symptoms Over 16 Weeks Post Intake

There were high levels of missingness in the PHQ-8 over the 16 weeks after the intake observation period. Of the 5890 clients, 3447 (58.5%) completed a 3-week survey, 1978 (33.6%) completed a 6-week survey, 1165 (19.8%) completed a 9-week survey, 724 (12.3%) completed a 12-week survey, and 506 (8.6%) completed a 15-week survey. When comparing those with at least one follow-up time point to those without, there were no differences in the rates of missing data on the basis of age, education level, history of therapy, gender, or PHQ-8 score.

Those who did not have any follow-up PHQ-8 were more likely to be divorced (8.1% with follow-up data vs 9.6% missing all follow-up) or living with a partner (8.1% vs 9.6%), less likely to be married (30.2% vs 27.1%), more likely to have a video plan (7.9% vs 9.4%), and less likely to have a messaging-only plan (90.6% vs 92.1%).

The best-fitting model for change over time in the PHQ-8 is shown in Figure 3, illustrating a curvilinear change in PHQ-8 scores during the first 16 weeks of treatment (quadratic $-2LL_5$=−41,432, deviance=399, P<.001), therefore all subsequent models included linear and quadratic change variables. The model-derived average PHQ-8 score at intake was 14.94 (SE 0.08), with a linear slope decreasing an average of −1.25 points (SE 0.03) between the first and second week and flattening by an additional 0.06 points (SE 0.002) for each subsequent week. Client scores on the PHQ-8 declined over time, dropping below the clinical cut-off from a score of 15 to 10 by week 6.

Table 5 displays models using single variables to predict change over time and found that client education level and age significantly predicted intake PHQ-8 score and linear change over time (education level linear – $2LL_2$ deviance=25, P<.001; age linear $2LL_6$ deviance=9, P=.006). Clients' PHQ-8 score at intake, and whether it was their first time in therapy, significantly predicted linear and quadratic change over time in the PHQ-8 score (first PHQ-8 score linear $-2LL_2$ deviance=1991, P<.001, quadratic $-2LL_1$ deviance=99, P<.001; first time in therapy linear $-2LL_2$ deviance=12, P<.001, quadratic $-2LL_1$ deviance=9, P<.001). Variables that were not significantly associated with PHQ-8 scores over time included plan type, client gender, therapist gender, therapist years of experience, and therapist endorsing depression-specific expertise. Model fit statistics are displayed in Table 6, demonstrating that, unsurprisingly, the model incorporating intake PHQ-8 score accounted for the largest amount of variance in the model.

The final model explained 51.1% of the variance in PHQ-8 score change. This model included predictor variables that had significant relationships with the intercept and linear or quadratic change in PHQ-8 scores over time, on the basis of single variable modeling (see Table 7). Significant predictors were as follows: those with an associate's degree or higher improved more quickly based on their sharper linear slope (linear estimate=−0.07, P=.002). Higher intake PHQ-8 scores, displayed as a median split in Figure 4, were associated with a larger PHQ-8 intercept (estimate=3.73, P<.001), faster linear improvement (estimate=−0.50, P<.001) and faster flattening (quadratic estimate=0.03, P<.001). There were no differences in the PHQ-8 intercept on whether it was the client's first time in therapy; however, first-time clients had faster linear improvement (linear estimate=−0.37, P<.001) and faster flattening (quadratic estimate=0.03, P<.001). Age was not significant after controlling for other variables.

Figure 3. Model estimated change on the Patient Health Questionnaire-8 item (PHQ-8) over 16 weeks with SE shading.

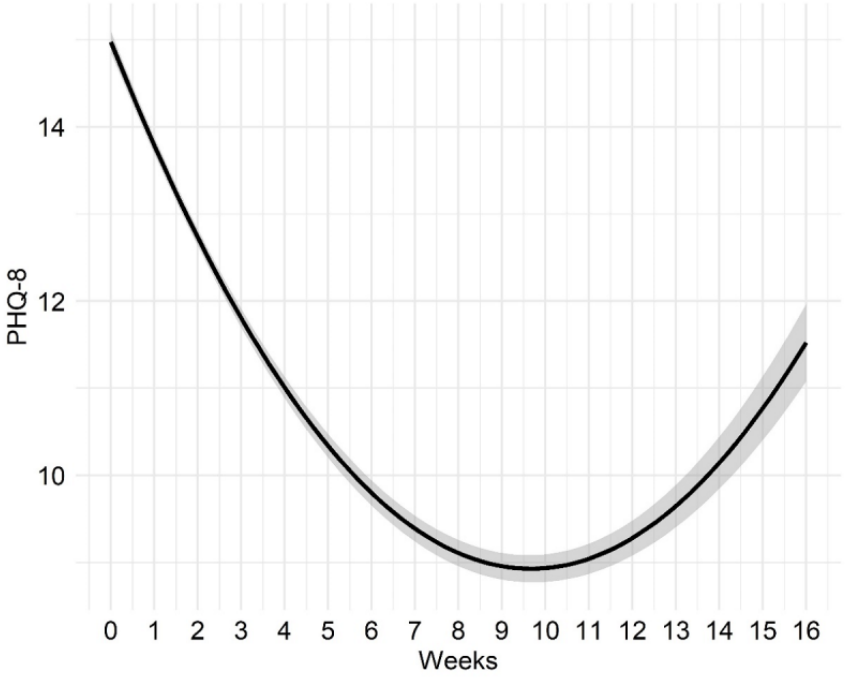


Table 5. Significant single-variable mixed effects models predicting Patient Health Questionnaire-8 item (PHQ-8) score change[a].

	Estimate	SE	*P* value
Education (ref=high school)			
Intercept	15.66	0.13	<.001
Week	−1.22	0.03	<.001
Week^2	0.06	0.002	<.001
Associate's or bachelor's degree or higher	−0.90	0.15	<.001
Associate's degree or higher × week	−0.02	0.02	.34
Age (ref=18-25 years)			
Intercept	15.26	0.13	<.001
Week	−1.24	0.03	<.001
Week^2	0.06	0.002	<.001
26-35	−0.49	0.14	<.001
36-49	−0.14	0.17	.41
>50	−0.42	0.29	.14
26-35 × week	−0.01	0.03	.67
36-49 × week	0.005	0.03	.87
>50 × week	0.05	0.05	.27
Intake PHQ-8 score			
Intercept	15.07	0.06	<.001
Week	−1.28	0.02	<.001
Week^2	0.07	0.002	<.001
PHQ-8	3.75	0.05	<.001
PHQ-8 × week	−0.49	0.02	<.001
PHQ-8 × week^2	0.03	0.002	<.001
First time in therapy			
Intercept	15.04	0.09	<.001
Week	−1.17	0.03	<.001
Week ^{2}	0.06	0.003	<.001
First time in therapy	−0.20	0.15	.17
First time in therapy × week	−0.30	0.07	<.001
First time in therapy × week^2	0.02	0.005	<.001

[a]Variables that were tested and were not significant on the basis of model fit statistics included client gender, marital status, plan type, therapist gender, therapist years of experience, and therapist endorsing depression-specific expertise.

Table 6. Model fit statistics predicting Patient Health Questionnaire-8 item (PHQ-8) score change.

	AIC[a]	BIC[b]	−2LL[c]	df	Conditional R^2	Marginal R^2
Individual models						
Education	66,153.2	66,204.5	−33,069.6	7	0.279	0.180
Age	66,082.0	66,162.7	−33,030	11	0.271	0.179
Intake PHQ-8 score	62,024.4	62,083.0	−31,004.2	8	0.510	0.460
First time in therapy	62,877.0	62,935.3	−31,430.5	8	0.278	0.182
Final combined model	64,112.8	64,252.3	−32,037.4	19	0.511	0.455

[a]AIC: Akaike Information Criterion.

[b]BIC: Bayesian Information Criterion.

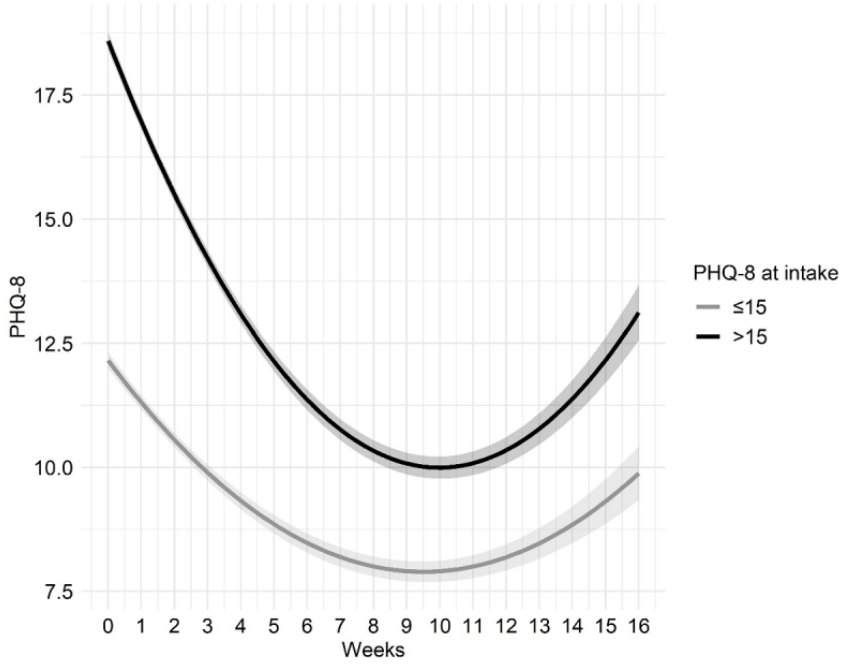
[c]−2LL: −2 Log Likelihood.

Table 7. Final combined mixed effects model predicting Patient Health Questionnaire (PHQ-8) score change[a].

	Estimate	SE	*P* value
Intercept	15.36	0.14	<.001
Week	−1.10	0.04	<.001
Week^2	0.06	0.002	<.001
Associate's or bachelor's degree or higher (ref=high school)	−0.18	0.12	.120
Associate's degree or higher × week (ref=high school)	−0.07	0.02	.002
Age (years, ref=18-25)			
26-35	−0.16	0.13	.20
36-49	−0.10	0.16	.52
>50	−0.41	0.26	.12
Age (years, ref=18-25) × week			
26-35	−0.03	0.02	.15
36-49	−0.01	0.03	.76
>50	0.04	0.04	.32
Intake PHQ-8 score	3.73	0.06	<.001
Intake PHQ-8 score × week	−0.50	0.03	<.001
Intake PHQ-8 score × week^2	0.03	0.002	<.001
First time in therapy	−0.07	0.13	.58
First time in therapy × week	−0.30	0.06	<.001
First time in therapy × week^2	0.03	0.005	<.001

[a]Predictor and predictor × time interaction variables that were significant in simpler models were included. The intake PHQ-8 scores were grand mean-centered.

Figure 4. Median split Patient Health Questionnaire-8 item (PHQ-8) score at intake and change over time with SE shading.



Discussion

Principal Findings

This is the first population-based study of treatment engagement, subsequent outcomes, and predictors of outcomes in the routine delivery of technology-mediated psychotherapy focused on adult clients experiencing clinically significant levels of depression. Our findings indicate that technology-mediated psychotherapy may be helpful in reducing symptoms of depression. On average, client scores on the PHQ-8 declined over time, dropping below the clinical cut-off from a score of 15 to 10 by week 6. At the same time point, 37% of clients had disengaged from the therapy. There were no differences in the length of treatment engagement or speed of improvement based on whether clients opted for a plan that included synchronous video chat (either once or 4 times per month) in addition to the asynchronous messaging plan.

Findings from our analyses predicting symptom change and disengagement suggest that the degree to which the therapy was helpful did not depend on the client or therapist characteristics included in the study; however, how quickly clients decided to end the therapy did, though this effect was very small. One exception was education level. Completing formal education beyond high school was related to a sharper decrease in PHQ-8 scores and a shorter length of time before disengagement, which may be interpreted as clients with higher education being a group that is particularly likely to benefit quickly from technology-mediated psychotherapy and, therefore, end it sooner. Consistent with this premise, there is some evidence that individuals with more years of formal education are less likely to end therapy prematurely [23]. Clients who were in therapy for the first time were also more likely to improve faster. While this might be associated with less severe depressive symptoms, this result was true even after controlling for baseline depression scores. It may be that those repeatedly seeking treatment have symptoms that are more chronic, even if they are of equal severity to those seeking treatment for the first time.

The fact that specific age categories and gender were not significantly associated with depressive symptom change is consistent with a study examining moderators of outcomes among primary care patients in the United Kingdom randomized to chat-based CBT therapy versus usual care [13]. However, in contrast to our findings, educational attainment did not moderate the relationship between therapy and outcomes in the primary care sample. This is also consistent with a recent meta-regression study examining moderators of outcomes from randomized trials of face-to-face in-person psychotherapy versus a control group among adults with depressive symptoms, in which the authors found no difference between older and younger adults and between men and women [24]. The authors note that there were insufficient studies available for the analysis with a low risk of bias to examine education level as a moderator.

Unfortunately, we did not have data on client race and ethnicity; however, in the meta-regression study mentioned [24], studies focusing on diverse samples did not have better or worse effect sizes than the other studies. This is similar to findings from a previous meta-analysis using a similar methodology, looking at randomized controlled trials in which race and ethnicity data were reported [25]. The authors examined whether the proportion of adults identified as a racial or ethnic minority in the sample was associated with treatment effect size across 56 trials and found no relation. Although psychotherapy appears to be effective across racial and ethnic groups [26], access to and barriers to engagement in psychotherapy are characterized by racial or ethnic disparities [27]. For instance, a large meta-analysis of randomized controlled trials of face-to-face in-person psychotherapy for depression among adults found that studies with a higher proportion of participants identifying with a racial or ethnic minority group had higher rates of early therapy termination, as defined by research teams [28].

In a nationally representative US-based study, people identified as Black or African American, Latinx, and Asian Americans were less likely to engage in mental health outpatient care and indicated financial costs to be a major barrier to accessing services [29]. Other barriers may be related to beliefs that therapy is not likely to be effective, stigma and shame about needing care and being seen accessing it, language barriers, and cultural practices that may not resonate with a Western medical model [27,30-32]. Technology-mediated psychotherapy holds promise for alleviating some of these barriers to psychotherapy, particularly in reducing the need to find a local therapist and attend in-person sessions that may require transportation and additional time to travel to appointments. Although technology-mediated therapy may increase a sense of privacy and confidentiality as clients do not have to present to a clinic or office for care where others may see them, there may be additional barriers to ensuring privacy that have to do with concern over security breaches [33]. Although technology and internet access are expanding rapidly, there are still substantial disparities in access to stable and consistent internet and internet-capable smartphones and computers [34]. Future research can examine the role of race, ethnicity, and the intersection of access to technology-mediated psychotherapy in both outcomes and therapeutic engagement.

Our findings also indicate that the length of engagement in services is weakly associated with the variables included in this study. More research should examine the determinants of engagement and work to operationalize them using feasible methods so that clinical care providers can target their efforts to prevent premature disengagement. While the association is weak, our findings indicate that being younger, in therapy for the first time, unmarried but living with their partner (as compared with single), and being with a therapist with less experience predicted a shorter length of time to disengagement from the therapy. Previous research indicates that younger clients are more likely to prematurely end therapy [23], and that less experienced therapists tend to have higher premature termination rates [23] and more difficulty engaging and retaining clients in therapy [35]. Marital status did not appear to be related to psychotherapy or mental health service disengagement in previous face-to-face psychotherapy studies [23,28,36]. However, a study of chat-based CBT with primary care patients observed that those patients grouped into a category of separated, divorced, or widowed had greater benefit from the therapy than those who were married or living as married and those who were single. The authors determined that the effect seemed to be largely owing to the separated, divorced, or widowed group remaining more symptomatic when waitlisted and obtaining usual care. Understanding the role of marital status in therapeutic gains is complicated by examination of this variable. Owing to the small sample size, categories that seem conceptually similar on the surface may collapse in analyses but may actually represent very distinct groups empirically [37]. For instance, being unmarried and living with a partner may hold a unique meaning for the millennial generation, which comprises a large proportion of our sample who are known to delay marriage [38], and combining these clients with married clients, as was done in the Button et al [13] study, could obscure findings that would emerge if examined separately, as was done in our analysis. Qualitative research may help elucidate the significance of living with a partner, which may be related to a more rapid disengagement from therapy.

Additional research is needed to understand why client sociodemographic factors may influence perceptions of therapy and decisions to end the therapy. However, it is important to note that not all therapeutic disengagement outcomes are negative. In fact, people are known to vary in how quickly their symptoms respond to psychotherapy, and there is evidence to suggest that people leave therapy when they feel they have a good enough level of change [39-41]. A previous study of Talkspace clients suggested that this is often the case. Specifically, in a study of clients experiencing depression or anxiety, one-third of those leaving therapy before 12 weeks of enrollment reported reasons for leaving, with 53% indicating satisfaction with reaching therapeutic goals as the reason [18]. Therefore, engagement in technology-mediated psychotherapy may require shorter engagement times than in-person session-based therapy, potentially because of the ability of clients to access therapists as needed, rather than the typical once a week or every other week schedule.

This study examined disengagement from the initial enrollment in therapy; however, it is possible that clients re-engaged, and there may be observable patterns of disengagement and re-engagement over time among the clientele. In addition, the concept of disengagement does not shed light on the nature of a client's engagement in therapy. Therapeutic engagement can be defined as all efforts made by clients during the course of treatment (both within and between sessions) toward the achievement of changes (treatment outcomes) [35]. Once a client has disengaged, it is clear that they no longer participate in the therapy; however, if the client is still in therapy, the degree to which they are genuinely engaged in the process with the therapist can vary substantially and is an important variable in treatment outcomes. In an earlier Talkspace study with depressed or anxious clients, greater amounts of communication between clients in therapists based on word counts across text, audio, and video messaging were related to symptom improvement [18]. Research on face-to-face in-person CBT in which the timing of sessions varied suggests that having more frequent sessions in closer proximity to each other versus longer courses of therapy is related to improved outcomes [42]. A benefit of technology-mediated psychotherapy, such as what the Talkspace platform offers, is that clients have more potential to increase the intensity of their therapy and perhaps more quickly benefit from it.

Interestingly, it may not be necessary for clients to meet with their therapist in a more traditional synchronous manner, such as having a face-to-face web-based video-based session, to receive an adequate degree of therapeutic intensity through technology-mediated therapy. In this study, we did not observe a relationship between the type of plan clients opted into and changes in depression scores over time, although some plan types included 1 or 4 synchronous video sessions per month in addition to text and audio or video messaging. In fact, the predominant type of plan clients opted for when they enrolled in the Talkspace platform was asynchronous messaging only

(n=5389, 91.5%), although we do not know whether clients opted out of the other plans owing to cost, preference, or both.

Limitations and Considerations

A key consideration for our findings is the nature of clients enrolled in technology-mediated psychotherapy. The most common client characteristics were being of the millennial generation [38], college educated, single, identifying as female, and having experience with therapy. In addition, as the messaging-only plan was by far the most frequently chosen (91.5%), the study reflects clients willing or preferring to engage in primarily text messaging–based therapy, which may or may not be asynchronous. Men are similarly underrepresented in face-to-face in-person therapy [43] and represent a population that may need additional effort to reach and engage in technology-mediated psychotherapy. The population in our study may represent what could be considered early adopters of the innovation of text-based psychotherapy, which is also consistent with the premise that members of the millennial generation are more generally engaged with internet technology and more comfortable using it in myriad ways [44]. In addition, data were collected before the COVID-19 pandemic, when engaging in technology-mediated health care was novel for most people. Considering social distancing policies to control the spread of COVID-19, technology-mediated health care proliferated, and the business of web-based therapy grew tremendously as the general demand for psychotherapy simultaneously increased [7,45]. The postpandemic population may be different from the population included in our study. In fact, a recent study published with postpandemic Talkspace clients suggests that anxiety is a much more common experience and comorbidity among clients since the pandemic started [46].

Our findings for depressive symptom change do not generalize well past the first 6 weeks of therapy, given that missing data on the PHQ-8 increases substantially after 6 weeks. A known contributing factor to missing data on the PHQ-8 is that clients stop taking the measure after disengaging from the therapy. Our estimates of later time points were based on the trajectory that was observed at the time the clients stopped taking the PHQ-8. Given that it is common for people with greater improvement in earlier phases of therapy to end therapy earlier than those whose symptoms do not remit as quickly [47], it is possible that the overall improvement in depression symptoms is an overestimation of what we would observe if clients continued taking PHQ-8 assessments. There are high rates of missing data, aside from that, which can be attributed to therapy disengagement. Therefore, these results may only be generalized to people who are receiving therapy and who complete the PHQ-8 assessments. In addition, given that we observed that higher intake PHQ-8 scores were predictive of symptom improvement, a likely caveat for our findings is that there is some regression to the mean, a statistical artifact of more extreme-scoring individuals to score more closely to the average on the next assessment. In randomized controlled trials, where regression to the mean is accounted for, it is more common to actually see higher initial depression scores as related to less symptom improvement [48-50]. Like all regression models, our curvilinear time modeling provides an overall average estimate of symptom trajectories and is not able to demonstrate what are likely nonlinear relationships with symptoms and time that would be observed if examined at the individual client level [51]. However, we can say that, on average, we see a trend for clients to do better symptomatically than when they enrolled.

We conducted a naturalistic study using routinely collected data from the Talkspace platform. Therefore, there were limitations in the amount and nature of the data available. First, we did not account for other conditions that clients may be seeking therapy for, or conditions that may moderate therapeutic outcomes. For instance, the presence of a personality disorder is related to premature termination of therapy [23] and is known to impede symptom improvement [52]. The PHQ-8 does not capture suicidal thinking, so we do not have estimates of suicide risk using the item from the PHQ-9 that asks about *thoughts that you would be better off dead, or of hurting yourself*, which is known to be predictive of suicidal behavior and a common question asked as part of routine suicide risk screening procedures [53]. In addition, we do not know what type of therapy is provided by licensed clinicians; however, therapy from various theoretical orientations is known to be effective for depression [54]. The availability of textual data allows for an ample exploration of what is happening in routine psychotherapy in this context. Future research can use the rich data made available through text messages and transcripts made from audio or video interactions to examine the degree to which evidence-based interventions and practices are part of the therapy and how this might be related to disengagement, engagement, and symptom outcomes.

Conclusions

Technology-mediated psychotherapy is becoming an increasingly popular alternative to in-person psychotherapy. Our findings add to the growing literature showing the benefits of technology-mediated psychotherapy in general and messaging-based therapy in particular. In addition, benefits were observed over a relatively brief period, with a reduction in symptoms observed within 16 weeks of therapy. Some populations may require additional efforts to remain engaged once they enroll, such as in younger adults. Future research is needed to examine how well technology-mediated psychotherapy addresses disparities in access to therapy and opportunities to enhance reach and ensure the effectiveness of this treatment modality across the full spectrum of adults experiencing depression.

Acknowledgments

This research was supported by the National Institute of Mental Health of the National Institutes of Health under award numbers K23MH118361, P50MH115837, and R44MH124334. The content is the sole responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Authors' Contributions

DD, MDP, TDH, and PA contributed to the conceptualization, review, and writing of the manuscript. DD, MDP, and SC formulated the data analytic approach. Data analysis was conducted using MDP and SC. DD, MDP, TDH, and PA drafted portions of the manuscript and provided substantive review of manuscript drafts. DD coordinated the manuscript preparation and revisions.

Conflicts of Interest

TDH is an employee of the service that provided the data.

References

1. Key Substance Use and Mental Health Indicators in the United States: Results from the 2019 National Survey on Drug Use and Health. Rockville, MD: Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration; 2020.
2. Kessler RC, Petukhova M, Sampson NA, Zaslavsky AM, Wittchen H. Twelve-month and lifetime prevalence and lifetime morbid risk of anxiety and mood disorders in the United States. Int J Methods Psychiatr Res 2012 Sep;21(3):169-184 [FREE Full text] [doi: 10.1002/mpr.1359] [Medline: 22865617]
3. Cuijpers P. Psychotherapies for adult depression: recent developments. Curr Opin Psychiatry 2015 Jan;28(1):24-29. [doi: 10.1097/YCO.0000000000000121] [Medline: 25415495]
4. McHugh RK, Whitton SW, Peckham AD, Welge JA, Otto MW. Patient preference for psychological vs pharmacologic treatment of psychiatric disorders. J Clin Psychiatry 2013 Jun 15;74(06):595-602. [doi: 10.4088/jcp.12r07757]
5. Mohr DC, Hart SL, Howard I, Julian L, Vella L, Catledge C, et al. Barriers to psychotherapy among depressed and nondepressed primary care patients. Ann Behav Med 2006 Dec;32(3):254-258 [FREE Full text] [doi: 10.1207/s15324796abm3203_12] [Medline: 17107299]
6. Imel ZE, Caperton DD, Tanana M, Atkins DC. Technology-enhanced human interaction in psychotherapy. J Couns Psychol 2017 Jul;64(4):385-393 [FREE Full text] [doi: 10.1037/cou0000213] [Medline: 28318277]
7. Steele A. Pandemic, new platforms prompt surge in new therapists. Wall Street Journal. 2020. URL: https://www.wsj.com/articles/pandemic-new-platforms-prompt-surge-in-new-therapists-11601816400 [accessed 2021-12-28]
8. Castro A, Gili M, Ricci-Cabello I, Roca M, Gilbody S, Perez-Ara MA, et al. Effectiveness and adherence of telephone-administered psychotherapy for depression: a systematic review and meta-analysis. J Affect Disord 2020 Jan 01;260:514-526. [doi: 10.1016/j.jad.2019.09.023] [Medline: 31539688]
9. Osenbach JE, O'Brien KM, Mishkind M, Smolenski DJ. Synchronous telehealth technologies in psychotherapy for depression: a meta-analysis. Depress Anxiety 2013 Nov 06;30(11):1058-1067. [doi: 10.1002/da.22165] [Medline: 23922191]
10. Mohr DC, Vella L, Hart S, Heckman T, Simon G. The effect of telephone-administered psychotherapy on symptoms of depression and attrition: a meta-analysis. Clin Psychol (New York) 2008;15(3):243-253 [FREE Full text] [doi: 10.1111/j.1468-2850.2008.00134.x] [Medline: 21369344]
11. Vernmark K, Lenndin J, Bjärehed J, Carlsson M, Karlsson J, Oberg J, et al. Internet administered guided self-help versus individualized e-mail therapy: a randomized trial of two versions of CBT for major depression. Behav Res Ther 2010 May;48(5):368-376. [doi: 10.1016/j.brat.2010.01.005] [Medline: 20152960]
12. Kessler D, Lewis G, Kaur S, Wiles N, King M, Weich S, et al. Therapist-delivered internet psychotherapy for depression in primary care: a randomised controlled trial. Lancet 2009 Aug;374(9690):628-634. [doi: 10.1016/s0140-6736(09)61257-5]
13. Button KS, Wiles NJ, Lewis G, Peters TJ, Kessler D. Factors associated with differential response to online cognitive behavioural therapy. Soc Psychiatry Psychiatr Epidemiol 2012 May 4;47(5):827-833. [doi: 10.1007/s00127-011-0389-1] [Medline: 21541696]
14. Barak A, Hen L, Boniel-Nissim M, Shapira N. A comprehensive review and a meta-analysis of the effectiveness of internet-based psychotherapeutic interventions. J Technol Human Service 2008 Jul 03;26(2-4):109-160. [doi: 10.1080/15228830802094429]
15. Nitzburg GC, Farber BA. Patterns of utilization and a case illustration of an interactive text-based psychotherapy delivery system. J Clin Psychol 2019 Feb 10;75(2):247-259. [doi: 10.1002/jclp.22718] [Medline: 30628062]
16. Hull TD, Mahan K. A study of asynchronous mobile-enabled SMS text psychotherapy. Telemed J E Health 2017 Mar;23(3):240-247. [doi: 10.1089/tmj.2016.0114] [Medline: 27797646]
17. DellaCrosse M, Mahan K, Hull TD. The effect of messaging therapy for depression and anxiety on employee productivity. J technol behav sci 2018 Jun 26;4(1):1-5. [doi: 10.1007/s41347-018-0064-4]
18. Hull TD, Malgaroli M, Connolly PS, Feuerstein S, Simon NM. Two-way messaging therapy for depression and anxiety: longitudinal response trajectories. BMC Psychiatry 2020 Jun 12;20(1):297 [FREE Full text] [doi: 10.1186/s12888-020-02721-x] [Medline: 32532225]
19. Posner K, Brown GK, Stanley B, Brent DA, Yershova KV, Oquendo MA, et al. The Columbia-Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. Am J Psychiatry 2011 Dec;168(12):1266-1277 [FREE Full text] [doi: 10.1176/appi.ajp.2011.10111704] [Medline: 22193671]

20. Kroenke K, Strine TW, Spitzer RL, Williams JB, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord 2009 Apr;114(1-3):163-173. [doi: 10.1016/j.jad.2008.06.026] [Medline: 18752852]
21. Löwe B, Unützer J, Callahan CM, Perkins AJ, Kroenke K. Monitoring depression treatment outcomes with the patient health questionnaire-9. Med Care 2004 Dec;42(12):1194-1201. [doi: 10.1097/00005650-200412000-00006] [Medline: 15550799]
22. Singer JD, Willett JB. Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence. Oxford, UK: Oxford Scholarship; 2003.
23. Swift JK, Greenberg RP. Premature discontinuation in adult psychotherapy: a meta-analysis. J Consult Clin Psychol 2012 Aug;80(4):547-559. [doi: 10.1037/a0028226] [Medline: 22506792]
24. Cuijpers P, Karyotaki E, Reijnders M, Huibers MJ. Who benefits from psychotherapies for adult depression? A meta-analytic update of the evidence. Cogn Behav Ther 2018 Mar 18;47(2):91-106. [doi: 10.1080/16506073.2017.1420098] [Medline: 29345530]
25. Ünlü Ince B, Riper H, van 't Hof E, Cuijpers P. The effects of psychotherapy on depression among racial-ethnic minority groups: a metaregression analysis. Psychiatr Serv 2014 May 01;65(5):612-617. [doi: 10.1176/appi.ps.201300165] [Medline: 24535615]
26. Huey SJ, Tilley JL, Jones EO, Smith CA. The contribution of cultural competence to evidence-based care for ethnically diverse populations. Annu Rev Clin Psychol 2014 Mar 28;10(1):305-338. [doi: 10.1146/annurev-clinpsy-032813-153729] [Medline: 24437436]
27. Tambling RR, Hynes KC, D'Aniello C. Are barriers to psychotherapy treatment seeking indicators of social determinants of health?: a critical review of the literature. Am J Family Ther 2021 Jan 08:1-16. [doi: 10.1080/01926187.2020.1867931]
28. Cooper AA, Conklin LR. Dropout from individual psychotherapy for major depression: a meta-analysis of randomized clinical trials. Clin Psychol Rev 2015 Aug;40:57-65. [doi: 10.1016/j.cpr.2015.05.001] [Medline: 26067572]
29. Racial/ethnic differences in mental health service use among adults. Substance Abuse and Mental Health Services Administration. URL: https://www.samhsa.gov/data/report/racialethnic-differences-mental-health-service-use-among-adults [accessed 2021-12-27]
30. Barrera I, Longoria D. Examining cultural mental health care barriers among Latinos. J Leadership Equity Res 2018;4(1).
31. Chen J, Rizzo J. Racial and ethnic disparities in use of psychotherapy: evidence from U.S. national survey data. Psychiatr Serv 2010 Apr;61(4):364-372. [doi: 10.1176/ps.2010.61.4.364] [Medline: 20360275]
32. Thompson VL, Bazile A, Akbar M. African Americans' perceptions of psychotherapy and psychotherapists. Professional Psychol Res Practice 2004;35(1):19-26. [doi: 10.1037/0735-7028.35.1.19]
33. Renn BN, Hoeft TJ, Lee HS, Bauer AM, Areán PA. Preference for in-person psychotherapy versus digital psychotherapy options for depression: survey of adults in the U.S. NPJ Digit Med 2019 Feb 11;2(1):6 [FREE Full text] [doi: 10.1038/s41746-019-0077-1] [Medline: 31304356]
34. Marler W. Mobile phones and inequality: findings, trends, and future directions. New Media Soc 2018 Apr 07;20(9):3498-3520. [doi: 10.1177/1461444818765154]
35. Holdsworth E, Bowen E, Brown S, Howat D. Client engagement in psychotherapeutic treatment and associations with client characteristics, therapist characteristics, and treatment factors. Clin Psychol Rev 2014 Jul;34(5):428-450. [doi: 10.1016/j.cpr.2014.06.004] [Medline: 25000204]
36. O'Brien A, Fahmy R, Singh SP. Disengagement from mental health services. A literature review. Soc Psychiatry Psychiatr Epidemiol 2009 Jul 26;44(7):558-568. [doi: 10.1007/s00127-008-0476-0] [Medline: 19037573]
37. Afifi TO, Cox BJ, Enns MW. Mental health profiles among married, never-married, and separated/divorced mothers in a nationally representative sample. Soc Psychiatry Psychiatr Epidemiol 2006 Feb 9;41(2):122-129. [doi: 10.1007/s00127-005-0005-3] [Medline: 16467954]
38. Millennial life: how young adulthood today compares with prior generations. Pew Research Center. URL: https://www.researchgate.net/publication/337113826_Millennial_life_How_young_adulthood_today_compares_with_prior_generations [accessed 2021-12-27]
39. Lee AA, Sripada RK, Hale AC, Ganoczy D, Trivedi RB, Arnow B, et al. Psychotherapy and depressive symptom trajectories among VA patients: comparing dose-effect and good-enough level models. J Consult Clin Psychol 2021 May;89(5):379-392. [doi: 10.1037/ccp0000645] [Medline: 34124925]
40. Simon GE, Imel ZE, Ludman EJ, Steinfeld BJ. Is dropout after a first psychotherapy visit always a bad outcome? Psychiatr Serv 2012 Jul;63(7):705-707 [FREE Full text] [doi: 10.1176/appi.ps.201100309] [Medline: 22752034]
41. Barkham M, Connell J, Stiles WB, Miles JN, Margison F, Evans C, et al. Dose-effect relations and responsive regulation of treatment duration: the good enough level. J Consult Clin Psychol 2006;74(1):160-167. [doi: 10.1037/0022-006x.74.1.160]
42. Bruijniks SJ, Lemmens LH, Hollon SD, Peeters FP, Cuijpers P, Arntz A, et al. The effects of once- versus twice-weekly sessions on psychotherapy outcomes in depressed patients. Br J Psychiatry 2020 Apr 07;216(4):222-230. [doi: 10.1192/bjp.2019.265] [Medline: 32029012]
43. Seidler ZE, Rice SM, Ogrodniczuk JS, Oliffe JL, Shaw JM, Dhillon HM. Men, masculinities, depression: implications for mental health services from a Delphi expert consensus study. Professional Psychol Res Pract 2019 Feb;50(1):51-61. [doi: 10.1037/pro0000220]

44. Vogels E. Millennials stand out for their technology use, but older generations also embrace digital life. Pew Research Center. URL: https://www.pewresearch.org/fact-tank/2019/09/09/us-generations-technology-use/ [accessed 2021-12-27]
45. Worsening mental health crisis pressures psychologist workforce COVID-19 Practitioner Survey. American Psychological Association. URL: https://www.apa.org/pubs/reports/practitioner/covid-19-2021 [accessed 2021-12-27]
46. Hull TD, Levine J, Bantilan N, Desai AN, Majumder MS. Analyzing digital evidence from a telemental health platform to assess complex psychological responses to the COVID-19 pandemic: content analysis of text messages. JMIR Form Res 2021 Feb 09;5(2):e26190 [FREE Full text] [doi: 10.2196/26190] [Medline: 33502999]
47. Niileksela CR, Ghosh A, Janis RA. The dose-effect and good enough level models of change for specific psychological concerns. J Consult Clin Psychol 2021 Mar;89(3):200-213. [doi: 10.1037/ccp0000635] [Medline: 33829808]
48. Elkin I, Gibbons RD, Shea MT, Sotsky SM, Watkins JT, Pilkonis PA, et al. Initial severity and differential treatment outcome in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. J Consult Clin Psychol 1995;63(5):841-847. [doi: 10.1037/0022-006x.63.5.841]
49. Delgadillo J, Moreea O, Lutz W. Different people respond differently to therapy: a demonstration using patient profiling and risk stratification. Behav Res Ther 2016 Apr;79:15-22. [doi: 10.1016/j.brat.2016.02.003] [Medline: 26937855]
50. Katon W, Unützer J, Russo J. Major depression: the importance of clinical characteristics and treatment response to prognosis. Depress Anxiety 2010 Jan;27(1):19-26. [doi: 10.1002/da.20613] [Medline: 19798766]
51. Helmich MA, Wichers M, Olthof M, Strunk G, Aas B, Aichhorn W, et al. Sudden gains in day-to-day change: revealing nonlinear patterns of individual improvement in depression. J Consult Clin Psychol 2020 Feb;88(2):119-127. [doi: 10.1037/ccp0000469] [Medline: 31894994]
52. Newton-Howes G, Tyrer P, Johnson T. Personality disorder and the outcome of depression: meta-analysis of published studies. Br J Psychiatry 2006 Jan 02;188(1):13-20. [doi: 10.1192/bjp.188.1.13] [Medline: 16388064]
53. Simon GE, Rutter CM, Peterson D, Oliver M, Whiteside U, Operskalski B, et al. Does response on the PHQ-9 Depression Questionnaire predict subsequent suicide attempt or suicide death? Psychiatr Serv 2013 Dec 01;64(12):1195-1202 [FREE Full text] [doi: 10.1176/appi.ps.201200587] [Medline: 24036589]
54. Cuijpers P, van Straten A, Andersson G, van Oppen P. Psychotherapy for depression in adults: a meta-analysis of comparative outcome studies. J Consult Clin Psychol 2008 Dec;76(6):909-922. [doi: 10.1037/a0013075] [Medline: 19045960]

Abbreviations

CBT: cognitive behavior therapy
OR: odds ratio
PHQ: Patient Health Questionnaire

Edited by A Mavragani; submitted 16.01.22; peer-reviewed by S Rostam Niakan Kalhori , A Tannoubi; comments to author 19.04.22; revised version received 06.05.22; accepted 22.05.22; published 22.06.22.

Please cite as:
Darnell D, Pullmann MD, Hull TD, Chen S, Areán P
Predictors of Disengagement and Symptom Improvement Among Adults With Depression Enrolled in Talkspace, a Technology-Mediated Psychotherapy Platform: Naturalistic Observational Study
JMIR Form Res 2022;6(6):e36521
URL: https://formative.jmir.org/2022/6/e36521
doi:10.2196/36521
PMID:35731563

Original Paper

Using the Consolidated Framework for Implementation Research to Inform the Design of the Mobile Inspección Visual con Ácido Acético System: Mixed Methods Case Study

Hadley Woodruff Reid[1], MHS, MD; Rae Jean Proeschold-Bell[2,3], PhD; Christina Makarushka[4], MPH; Katherine Dayllan Melgar Vega[5], Lic; Megan Huchko[1,2,6], MPH, MD; Jose Jeronimo[5], MD; Lavanya Vasudevan[1,2,3,4], MPH, PhD

[1]Duke University School of Medicine, Durham, NC, United States

[2]Duke Global Health Institute, Durham, NC, United States

[3]Center for Health Policy and Inequalities Research, Duke University, Durham, NC, United States

[4]Department of Family Medicine and Community Health, Duke University, Durham, NC, United States

[5]La Liga Contra el Cancer, Lima, Peru

[6]Department of Obstetrics and Gynecology, Duke University, Durham, NC, United States

Corresponding Author:
Lavanya Vasudevan, MPH, PhD
Department of Family Medicine and Community Health
Duke University
Suite 600
2200 West Main Street
Durham, NC, 27705
United States
Phone: 1 919 613 1423
Email: lavanya.vasudevan@duke.edu

Abstract

Background: There is growing evidence supporting the use of mobile health (mHealth) interventions in low- and middle-income countries to address resource limitations in the delivery of health information and services to vulnerable populations. In parallel, there is an increasing emphasis on the use of implementation science tools and frameworks for the early identification of implementation barriers and to improve the acceptability, appropriateness, and adoption of mHealth interventions in resource-limited settings. However, there are limited examples of the application of implementation science tools and frameworks to the formative phase of mHealth design for resource-limited settings despite the potential benefits of this work for enhancing subsequent implementation, scale-up, and sustainability.

Objective: We presented a case study on the use of an implementation science framework in mHealth design. In particular, we illustrated the usability of the Consolidated Framework for Implementation Research (CFIR) for organizing and interpreting formative research findings during the design of the mobile Inspección Visual con Ácido Acético (mIVAA) system in Lima, Peru.

Methods: We collected formative data from prospective users of the mIVAA intervention using multiple research methodologies, including structured observations, surveys, group and individual interviews, and discussions with local stakeholders at the partnering organization in Peru. These activities enabled the documentation of clinical workflows, perceived barriers to and facilitators of mIVAA, overarching barriers to cervical cancer screening in community-based settings, and related local policies and guidelines in health care. Using a convergent mixed methods analytic approach and the CFIR as an organizing framework, we mapped formative research findings to identify key implementation barriers and inform iterations of the mIVAA system design.

Results: In the setting of our case study, most implementation barriers were identified in the CFIR domains of intervention characteristics and inner setting. All but one barrier were addressed before mIVAA deployment by modifying the system design and adding supportive resources. Solutions involved improvements to infrastructure, including cellular data plans to avoid disruption from internet failure; improved process and flow, including an updated software interface; and better user role definition for image capture to be consistent with local health care laws.

Conclusions: The CFIR can serve as a comprehensive framework for organizing formative research data and identifying key implementation barriers during mHealth intervention design. In our case study of the mIVAA system in Peru, formative research contributing to the CFIR domains of intervention characteristics and inner setting elicited the most key barriers to implementation. The early identification of barriers enabled design iterations before system deployment. Future efforts to develop mHealth interventions for low- and middle-income countries may benefit from using the approach presented in this case study as well as prioritizing the CFIR domains of intervention characteristics and inner setting.

(JMIR Form Res 2022;6(6):e32577) doi:10.2196/32577

KEYWORDS

cervical cancer; mobile health; Peru; colposcopy; implementation science; Consolidated Framework for Implementation Research; CFIR

Introduction

Background

With the increasing number of mobile phone and data subscribers worldwide, mobile health (mHealth; ie, the use of mobile technologies for delivering health services and information) has become a global phenomenon [1,2]. In many low- and middle-income countries (LMICs), mHealth interventions have been successfully used to mitigate health system challenges, including human resource, infrastructure, and information constraints [2-4]. Despite the multitude of mHealth interventions that have been piloted in LMICs, few have ultimately been brought to scale or had lasting sustainability [5]. The reasons for this leaky pilot-to-scale pipeline are varied, but a key reason may be the failure to identify and address implementation barriers, especially during the early stages of mHealth intervention development [6,7].

Strategies to support the effective translation of evidence-based interventions to real-world settings fall within the purview of implementation science [8]. Implementation science tools and frameworks can guide the exploration of implementation factors, facilitate the contextualization of those factors, identify evaluation metrics and benchmarks, and provide clues to enable intervention scale-up and sustainability [9]. In recognition of the benefits of implementation science in the context of scaling mHealth interventions, the World Health Organization and others have published several guiding documents that incorporate implementation science principles and methodologies in mHealth intervention design, evaluation, and reporting [3,10,11]. These guiding documents emphasize best practices such as early stakeholder engagement; needs assessment; contextual adaptation of intervention components; interoperability with extant systems; and assessment of process outcomes such as intervention acceptability, fidelity, and adoption to explain intervention effectiveness or a lack thereof. However, in the current research paradigm, the assessment of implementation factors typically occurs concurrently with or after the evaluation of mHealth intervention effectiveness (eg, using hybrid effectiveness-implementation study designs) [12]. As a result, there are limited illustrations of how implementation science tools and frameworks may be used to pre-empt barriers, inform mHealth intervention design, and increase the chances of successful implementation and scale-up [13].

The formative phase of mHealth intervention design provides a novel opportunity to assess any unique technical or practical considerations that may influence implementation. These considerations may include acceptability of mHealth interventions in an LMIC setting, mobile phone literacy of end users, and the cultural appropriateness of intervention components. In addition, data may be needed on the feasibility of stand-alone mHealth interventions or the organizational paradigm shifts that are needed to support integration of mHealth into the health system, clinical workflows, and existing data collection and management mechanisms [14]. Hence, conducting formative research to thoroughly understand the implementation environment, barriers, and facilitators is critical for informing mHealth implementation, scale-up, and sustainability. Despite this need, existing implementation science tools and frameworks are rarely applied to the formative phase of mHealth design.

In this paper, we describe the formative research phase for an mHealth intervention—namely, the mobile Inspección Visual con Ácido Acético (mIVAA) system for cervical cancer screening—as a case study on the use of implementation science frameworks to identify implementation barriers and inform intervention design. We present the utility of the Consolidated Framework for Implementation Research (CFIR) as a convergent framework to organize formative research findings and highlight key implementation barriers. We further motivate the use of such an approach to elucidate and mitigate potential stumbling blocks for the success of mHealth implementation, which can be addressed before investment in pilot studies.

The Problem and Proposed Digital Health Intervention

In Peru, cervical cancer is a significant contributor to mortality and morbidity among women of reproductive age. The incidence of cervical cancer in Peru is nearly twice the global rate (23.2 compared with 13.1 per 100,000 women) [15,16]. Evidence-based strategies such as early screening and preventative treatment can reduce morbidity and mortality from the disease [17]. Effective screening tests such as the Papanicolaou smear and, more recently, human papillomavirus DNA testing have successfully reduced the incidence of and mortality from cervical cancer in the United States and other high-income countries [18-20]. However, many of these strategies require resources (eg, laboratories) or personnel for implementation, and their widespread use in community-based prevention programs may be limited in LMIC settings. Thus, the World Health Organization recommends visual inspection with acetic acid (VIA) in situations where the capacity for human papillomavirus testing is lacking [17]. In VIA, health

workers examine the cervix with the naked eye for aceto-whitening, which is a sign of precancerous lesions.

La Liga Peruana Contra el Cáncer (La Liga) is a nonprofit organization based in Lima, Peru. La Liga routinely organizes free large-scale screening campaigns using both Papanicolau smears and VIA with community outreach mobile units that travel around the Lima metropolitan area to low-income neighborhoods. However, the slow turnaround of Papanicolau smear results and lack of patient access to reliable means of transportation mean that approximately 77% of screened-positive women in community settings are lost to subsequent clinic-based follow-up. In consultation with key decision makers at La Liga, the use of teleconsultation with a mobile phone–based software platform combined with visual counseling using a patient's cervical images was identified as a potential strategy to reduce this loss to follow-up. We hypothesized that this strategy may decrease loss to follow-up by using images of patients' own anatomy to reinforce the need for clinic-based follow-up among women with suspected cancer or precancer and by increasing access to early preliminary diagnosis through expert feedback when available. Hence, La Liga partnered with Duke University and a Peruvian partner, Medical Innovation and Technology, for the design of the mIVAA system. The mIVAA system is a mobile phone–based telemedicine platform that permits the documentation of magnified cervical images for remote and asynchronous expert colposcopist feedback. Imaging of the cervix can be achieved with the built-in mobile phone camera or through a USB-connected digital colposcopy device such as a pocket colposcope [21,22]. Formative research was conducted to inform the design and implementation of the mIVAA system in La Liga's community outreach units with the goal of identifying and mitigating implementation barriers to the maximum extent possible during the intervention development and pilot.

Theoretical Framework

The CFIR is a framework of implementation factors often used to assess readiness for implementation and comprises 5 overarching domains: intervention characteristics, inner setting, outer setting, characteristics of individuals, and process [23]. CFIR constructs have been shown to map well to common implementation challenges in the use of digital technologies [13,24]. The CFIR has been used to evaluate the implementation of mHealth interventions in low-resource contexts, including for longitudinal assessment of 2-way SMS interventions on HIV therapy adherence and an app tracking communicable disease transmission in Uganda [25,26]. However, in both of these studies, the CFIR was used after the rollout of the mHealth intervention. The CFIR can be applied earlier in the development of interventions; however, there are limited examples of the use of this framework to guide mHealth intervention development in formative research.

We selected the CFIR for our case study as it is a well-studied and widely used framework for implementation research. Furthermore, it contains a comprehensive list of implementation factors that can be evaluated by researchers, including those with a limited background in implementation science. Finally, it also permits the description of factors related to both the general implementation context and intervention-specific factors.

Our findings demonstrate the utility of the CFIR for organizing implementation factors emerging during the formative phase of mHealth design as well as for guiding mHealth intervention development and implementation.

Research Aims

The goal of this case study was to describe the use of the CFIR to categorize facilitators of and barriers to implementation identified during the design of the mIVAA system for cervical cancer screening in Lima, Peru. The secondary aim was to use the CFIR to inform solutions to the identified barriers before the pilot implementation of the mIVAA system.

Methods

Reporting

This study is reported in accordance with the consensus standards for the reporting of case studies (Table S1 in Multimedia Appendix 1) [27].

Study Setting

This study was conducted in partnership with La Liga in Lima, Peru. The mIVAA system was designed to be used in La Liga's community outreach units, which are staffed by a midwife (*obstetra*) and nurse technician (*técnica*). The mIVAA system comprises a digital imaging device and a telemedicine platform. The system can be used by midwives to acquire cervical images of patients in community settings and share them with expert colposcopists based in La Liga's brick-and-mortar clinics for feedback to inform triage options. Midwives concurrently perform Papanicolau smears; however, the mIVAA system augments the naked-eye visual examination and allows for documentation via imaging.

Study Design, Systems, and Frameworks

From April 2019 to October 2019, we conducted formative research using mixed methods to gather data on prospective implementation factors that could affect the design of the mIVAA system for cervical cancer screening. We selected the CFIR post hoc as an analytic framework for organizing the study findings for the aforementioned reasons.

Participants

The participants consisted of health care providers and staff who were involved in the cervical cancer screening workflow at La Liga, including midwives and nurse technicians at La Liga's 5 mobile community outreach units, colposcopists at the La Liga brick-and-mortar clinics, staff involved in patient follow-up and appointment scheduling, and La Liga administrators with decision-making authority. All eligible participants who were approached consented to take part in the study.

The same staff members who were observed in their workflow were approached for subsequent interviews, and all staff members were invited to participate in a group discussion and survey. The participants were involved in all parts of data

collection as each type of data collection had a specific purpose. The observations allowed the study team to inspect the workflow in detail; the interviews solicited individuals' thoughts on the mIVAA system and implementation context; the survey elicited general attitudes and readiness for implementation at La Liga; and, finally, the group discussions allowed for the synthesis of findings, feedback on proposed solutions, and elicitation of any residual barriers.

Ethics Approval

This study was approved by the Institutional Review Boards of the Duke University Health System (protocol Pro00102194) and by the University of San Martín De Porres (092-2019) in Peru. All participants provided informed consent before taking part in the study and were informed of the risks and benefits of participation as well as their right to stop participating at any time.

Data Collection

We used qualitative and quantitative methods to collect data, as described in the following sections.

Qualitative Data

Observations

Observations (n=18) of routine mobile unit and clinic workflows were conducted and recorded using a semistructured guide that emphasized documenting workflow, data collection systems, and opportunities for the integration of an mHealth telecolposcopy system with existing workflows and systems.

Individual Interviews and Group Discussions

Qualitative data were collected from midwives and nurse technicians (9/20, 45%) who staffed the mobile community outreach units and conducted cervical cancer screening, colposcopists (4/20, 20%) who performed diagnostic work at the La Liga brick-and-mortar clinics, staff who assisted with patient follow-up (2/20, 10%), and clinic-based administrators (5/20, 25%). Interview guides focused on the workflow of cancer screening appointments at mobile units, factors that might facilitate or inhibit the effective implementation of mHealth-supported telecolposcopy, and additional advice on incorporating telecolposcopy screening into usual clinical activities.

Written observation guides were translated from Spanish into English by bilingual study staff. Audio-recorded interviews and group discussions were first transcribed in Spanish and then translated into English by bilingual study staff. Additional bilingual study staff then read through the translated transcripts for grammar and logical flow while comparing the translations with the original Spanish.

Quantitative Data

We administered a cross-sectional survey to 22 participants, including midwives, nurse technicians, colposcopists, and administrative staff members. The survey included open-ended questions on the barriers to and facilitators of women receiving screening and treatment for cervical cancer, as well as the 15-item Evidence-based Practice Attitude Scale (EBPAS) examining attitudes toward the treatments, interventions, and systems related to cervical cancer [28]. The detailed methods and results of this survey have been reported previously [29].

Other Data

We also drew on internal organization documents and reports shared by our study partner, unpublished data from conversations with La Liga leadership, La Liga organizational materials such as descriptions of the organization mission and structure, and Peruvian telehealth legislation.

Data Analysis

Qualitative Data Analysis

Qualitative data analysis occurred in 2 stages.

Rapid Analysis

We used rapid analysis techniques that included a systems perspective, triangulation of data, additional data collection and input from study participants (who, as noted previously, were clinic providers and staff), and an iterative process that included decision-making with study participants [30]. In the first step of the rapid analysis, members of the study team (LV, CM, and KDMV) reviewed observation notes and interview transcripts as they were collected and developed actionable outputs, including workflow diagrams, mIVAA system mock-ups, and questions for gathering feedback from study participants. Using consensus discussions with JJ and other key decision makers at La Liga, the study team arrived at potential system features that were illustrated in the mock-ups. In the second step of the rapid analysis, the study team presented the outputs to the study participants during group discussions to elicit feedback and any residual concerns or barriers related to mIVAA implementation. This feedback was used during subsequent discussions to refine the system design, and additional potential solutions were brainstormed with the study team members at La Liga and Medical Innovation and Technology to determine a course of action for system development and implementation. The rapid analysis approach was necessary to facilitate progress toward system development and pilot evaluation within the time frame of the study funding [31].

Content Analysis

To confirm and further elaborate on the findings from the rapid analysis approach, we conducted formal content analysis on all qualitative data, including the observation notes, individual interviews, and group discussion transcripts, in parallel to system development and pilot implementation [32]. At this stage, the CFIR framework was chosen as the organizing structure for all data. A member of the research team, HWR, conducted an initial data-driven content analysis looking for mentions of potential barriers to the implementation of mHealth interventions. She documented each potential barrier, any proposed solution, and all representative quotes for each barrier. In addition, she categorized each barrier under the relevant CFIR construct using a table in Microsoft Word. The barriers and CFIR categorization were subsequently reviewed with 2 other study team members, LV and RJP-B, and each potential barrier was discussed until consensus was reached for categorization under a CFIR construct. In addition, RJP-B evaluated the quotations, and a single representative quote was selected for each barrier. These

findings were triangulated with the quantitative data as described in the following sections.

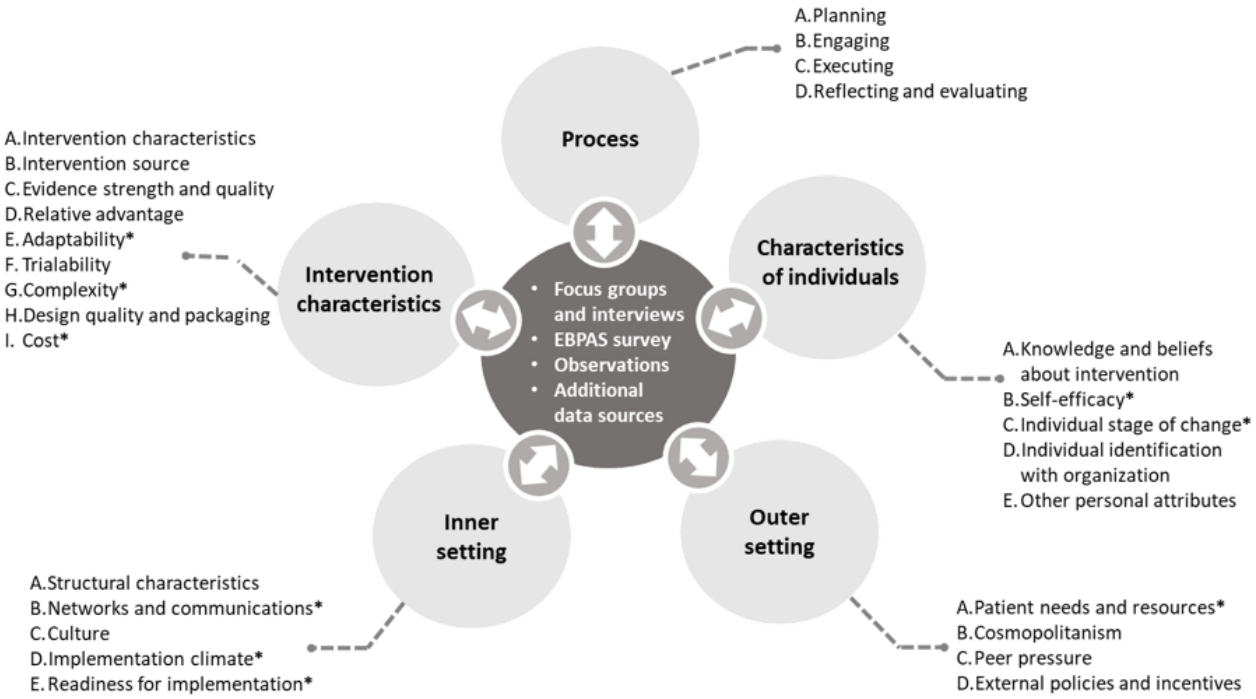
Quantitative Data Analysis

Summary statistics for the responses to the EBPAS were tabulated to quantify the implementation climate at La Liga and provider or staff willingness to adopt new strategies and interventions related to cervical cancer. We previously reported these findings on the EBPAS as well as other survey domains related to provider perceptions of patient-side barriers to screening and follow-up that are not included in this paper. Further details on the statistical methodology are described there [29].

Convergent Mixed Methods Analysis (Triangulation)

We used both qualitative and quantitative data from our formative work to assess each construct within the CFIR domains (Figure 1). The process domain constructs will be assessed in greater detail in a subsequent pilot implementation study (see Table S2 in Multimedia Appendix 1 for a full list of domains, descriptions, and data sources). Before interpretation, we mapped qualitative and quantitative data to each CFIR construct, including facilitators of implementation as well as descriptive data on the implementation context (Table S2 in Multimedia Appendix 1). In addition, we evaluated each barrier we identified through our content analysis and mapped it to its corresponding construct in the CFIR framework (see the asterisks in Figure 1). We triangulated the findings across data sources and tallied the number of barriers identified under each construct to identify the most salient constructs for successful pilot implementation of the mIVAA system. Data interpretation is presented in a narrative format in the *Results* and *Discussion* sections of this manuscript.

Figure 1. Data sources and integration of study findings using the Consolidated Framework for Implementation Research as a convergent database. *Constructs within which we identified barriers to implementation. EBPAS: Evidence-based Practice Attitude Scale.



Results

Facilitators of Implementation

Facilitators of implementation mapped broadly across the CFIR domains (Table S2 in Multimedia Appendix 1). These included many characteristics of our implementation setting—urban Peru—and our partner organization—La Liga—as well as the attitudes and perceptions of the individual La Liga staff members who would be implementing the mIVAA system. We found a receptive political environment within which La Liga was well situated to navigate local and national partnerships (outer setting). In addition, our survey data among La Liga staff showed a high willingness to accept new technologies related to cervical cancer (characteristics of individuals) [29]. However, overall, we chose to focus on barriers to implementation as these were critical to address before moving forward with the implementation of the mIVAA system.

Barriers to Implementation

Potential barriers to implementation in our study mapped to 4 of the 5 CFIR domains; namely, intervention characteristics, outer setting, inner setting, and characteristics of individuals (Table 1). Most identified barriers fell within the domain of intervention characteristics, including the constructs of adaptability (eg, internet connectivity), complexity (eg, role definition), and cost (eg, compensation for colposcopists' time). In addition, there were 5 potential barriers related to the domain of inner setting, with most of these related to how colposcopists would receive and process images captured with the mIVAA system and the resulting referrals. Other potential barriers included pragmatic considerations such as availability of electricity in mobile units to power digital devices and the

potential impact of vibrations in the mobile units on image quality, process concerns such as time to clean and sterilize the USB-connected imaging devices between patients, scheduling an increased number of follow-up appointments in La Liga's electronic health record system, and financial implications of the reduced number of patients able to be screened each day with the mIVAA system.

We identified 4 potential barriers to implementation that were not specific to the mIVAA system (Table S3 in Multimedia Appendix 1). These included time to sterilize speculums between clinic days (compatibility), the cost and distance of follow-up colposcopy for patients (patient needs and resources), incorrect patient phone numbers in the electronic health record system (patient needs and resources), and staff tardiness and presence at mobile units (other personal attributes). As these barriers may still affect the successful implementation of the mIVAA system, they informed data collection instruments for the pilot study.

Table 1. Barriers to mobile Inspección Visual con Ácido Acético (mIVAA) system implementation mapped to the Consolidated Framework for Implementation Research (CFIR) domains.

CFIR domain, construct, and potential barrier to implementation	Supporting quote or data notes
Intervention characteristics	
Adaptability	
Internet connectivity	"This is a system that depends on internet connectivity, it will be as good as the internet connection we have."
Vibrations and dust	"Some mobile units park on loosely packed earth and photo quality is affected by vibrations from movement in the unit."
Lack of electricity in some units	"Sometimes there is no light [electricity] or water, which delays the activities, because without light the tablets or the laptops wouldn't work."
Complexity	
Time for sterilization of USB-connected imaging device	"When you prepare in the squirt bottle [...] is a time that must be considered."
Conflicting priorities and timing of transmitting mIVAA images	"Most of the patients came to the mobile unit at 12 noon. This makes it difficult for the midwife to send the images to the colposcopists, complete patient's report, and other activities as soon as possible (prior to the closing time)."
Distinguishing results given after VIA[a] with mIVAA from Papanicolau results	"There in that moment [with images from the digital device] you are not going to tell her, lady look there isn't anything, come in a year. No, she [still] has to get her Pap [result] which was already taken."
Cost	
Financial impact of screening fewer women	"The goal of patients seen per day is 30 patients [...] With the implementation of the [mIVAA] we would have to evaluate how much the number of patients that are attended per day would decrease."
Cost of colposcopists' time	"We have to think about the budget...Right now I think it is a lie to say that a doctor will stop whatever he is doing to look at the screen and make that his priority."
Outer setting	
Patient needs and resources	
Delivery of results from mIVAA to women who have been screened	"...most of our women who are screened in the mobile units are mothers and have many duties at home, which can make it difficult for them to wait or return for their results."
External policy and incentives	
Legally allowable health care provision	Peruvian law states that only colposcopists may provide final review of cervical images, not midwives.
Inner setting	
Networks and communications	
Communication with and availability of colposcopists	"When the colposcopist is not at La Liga, we would have to find a way in which he would be able to connect to Wi-Fi and be able to review the images and send them back to the mobile unit."
Implementation climate-relative priority	
Colposcopists' desire for Papanicolau smear cytology results before evaluation	"In case the patient is scheduled for a colposcopy, we need to get Pap results as soon as possible, since some colposcopists only do colposcopy if patients have an alteration in their Pap results."
Implementation climate—compatibility	
Ability to correct a misdiagnosis on the telehealth platform	"What happens if I close it and maybe I made a mistake and I want to correct it?"
Ability to record mIVAA result in the existing WebLiga system	"Is there going to be a way to link this platform with the WebLiga system? Because when the study is done there must be a register of something."
Readiness for implementation—available resources	
MU[b] space for disinfecting baths	"Mobile Unit 4 has a small space only to put the laptop [not reprocessing baths]."

CFIR domain, construct, and potential barrier to implementation	Supporting quote or data notes
Scheduling follow-up appointments based on mIVAA results	"There might be an accumulation of patients to follow up, which can generate more workload."
Characteristics of individuals-self-efficacy	
Possibility of colposcopist missing a transmitted cervical image	"The oncologist gynecologist works in different institutions. At La Liga, they only work two to three times a week with medical appointments of two hours, so I believe that it is necessary to verify the read receipt of the images."

[a]VIA: visual inspection with acetic acid.

[b]MU: mobile unit.

Solutions for Implementation

The identified solutions for implementation varied (Table 2). In some cases, increased investment in infrastructure was necessary, such as providing cellular data plans to offset instances of poor internet connectivity or ensuring availability of the larger mobile units with adequate space for device setup and sterilization. Other barriers related to implementing mIVAA were able to be addressed in the subsequent pilot implementation protocol through improvements to the software interface and better role definition, such as assigning which staff member would transmit the cervical images throughout the day. A potential barrier, the ability to edit responses on the mIVAA software, could not be addressed in the pilot study because of budgetary limitations. An additional concern was that patients participating in the mIVAA intervention might mistake the results from the mIVAA screening with the final results from Papanicolau smear cytology. We did not directly address this as participants in the mIVAA intervention received the exact same counseling regarding Papanicolau smear results as women not participating in the intervention; therefore, there should not be a difference in understanding of the importance of Papanicolau smear cytology results.

Table 2. Solutions to identified implementation barriers by Consolidated Framework for Implementation Research (CFIR) domain.

CFIR domain, construct, and potential barrier to implementation	Solution or justification for not addressing it
Intervention characteristics	
Adaptability	
Internet connectivity	Providing phones with cellular plans to minimize reliance on internet connection in the mobile community outreach units. In addition, the mIVAA[a] app can be used offline to collect data; however, image transmission to colposcopists requires network connectivity.
Vibrations and dust	Providing tripods to enhance camera stability. Discouraging entry or exit of mobile unit while a photo is being taken to minimize vibrations.
Lack of electricity in some units	Providing an external phone battery pack. Phones can be used as a light source and provide power for the imaging device via USB.
Complexity	
Time for sterilization of USB-connected imaging device	Providing more than one imaging device (pocket colposcopes) per mobile unit to alternate between sterilization and use. Ability to use cell phone camera for image acquisition in the event that the pocket colposcopes are not ready for use.
Conflicting priorities and timing of transmitting mIVAA images	The midwife is asked to transmit images during the wait time between patients. To streamline data entry, the user interface of mIVAA is designed to be similar to the WebLiga system, and redundancy in data entry is minimized by using pictures of paper records.
Distinguishing results given after VIA[b] with mIVAA from Papanicolau results	Midwives continue to provide usual information to women on how to collect Papanicolau results.
Cost	
Financial impact of screening fewer women	Communicating financial impact to La Liga leaders and obtaining buy-in for lowering target recruitment to 20 patients per day during the pilot study.
Cost of colposcopists' time	Identifying and recruiting colposcopists willing to participate in the study with compensation provided for time spent reviewing study images.
Outer setting	
Patient needs and resources	
Delivery of results from mIVAA to women who have been screened	Adding a WhatsApp notification to colposcopists when new records are available for review to allow for same-day turnaround of results by the midwife. Colposcopists review patient records using a mobile app on their personal phone, which typically takes 2 to 3 minutes per patient.
External policy and incentives	
Legally allowable health care provision	Ensuring that midwife role is consistent with Peruvian guidelines and only colposcopists provide image review and diagnosis.
Inner setting	
Networks and communications	
Communication with and availability of colposcopists	Adding a WhatsApp notification when new records are available for review and allowing for review of patient records using a mobile app on their personal phone.
Implementation climate-relative priority	
Colposcopists' desire of Papanicolau results before evaluation	Working with La Liga decision makers to allow for prioritization of study participants presenting to colposcopy in laboratory queue for assessment of Papanicolau smears.
Implementation climate—compatibility	
Ability to correct a misdiagnosis on the telehealth platform	Not addressed in the current iteration of the mIVAA system because of budgetary limitations.
Ability to record mIVAA result in the existing WebLiga system	Incorporating the ability to generate a printout of the mIVAA report so it can be included in the paper medical record for each patient. La Liga is exploring options for direct data import into the WebLiga system.
Readiness for implementation—available resources	

CFIR domain, construct, and potential barrier to implementation	Solution or justification for not addressing it
Mobile unit space for disinfecting baths	Identifying space in the smaller mobile community outreach unit (eg, in the closet) that could be repurposed as space for disinfecting baths.
Scheduling increased number of follow-up appointments based on mIVAA results	Designing workflow for scheduling follow-up appointments with La Liga's administrative leadership and staff.
Characteristics of individuals-self-efficacy	
Possibility of colposcopist missing a transmitted cervical image	New records pending review are added to a common list allowing any colposcopist to claim and review the record. If the colposcopist does not review within 10 minutes of opening a record, the record is returned to the common list allowing other colposcopists to staff the case.

[a]mIVAA: mobile Inspección Visual con Ácido Acético.

[b]VIA: visual inspection with acetic acid.

Discussion

Principal Findings

Our study describes the use of the CFIR framework to identify and organize implementation barriers in the mHealth design phase. In our case study, formative research activities informed changes to system design based on identified barriers before piloting and evaluation. We used a mixed methods analytic approach relying on both quantitative and qualitative methods in this formative research phase. The survey data indicated that the participants were favorable to the implementation of new technology related to cervical cancer screening. The individual interview, group interview, and clinic flow observation data were triangulated as, in many cases, they pointed to the same barriers, such as the need for clear role delineation when transmitting images from the mIVAA system. However, in a few cases, a data source surfaced a unique barrier. For example, it was only through the clinic observations that we discovered the need to stabilize the device to prevent shaking from mobile clinics being parked on loosely packed earth. Using a framework such as the CFIR to organize this information is valuable for categorizing potential implementation barriers and facilitators. Previously, Westgard and Fleming [33] explored the use of the Active Implementation Frameworks to guide the design and implementation of an mHealth system for monitoring child health in the Amazon region of Peru. Although the Active Implementation Frameworks provide a comprehensive combined framework, their classification into 5 separate iterative frameworks can make them difficult to apply, especially for teams without formal implementation science expertise or limited resources for formative research. In addition, we did not use a technology-specific framework such as the technology acceptance model or the unified theory of acceptance and use of technology as we were interested in the broader implementation context and how to best integrate this intervention into the organizational structure, including external and internal influences.

The CFIR has been shown to map well onto the main challenges and findings of eHealth and mHealth research [13,25]. It was designed for use in formative research; however, assessment of the use of the CFIR has shown that this tool has been mainly applied to studies already in the pilot phase of implementation or beyond [24,26,34-36]. In the rare instances where it has been used before the pilot of an intervention, it has been useful in identifying and addressing barriers to implementation in an efficient manner [37]. We selected the CFIR post hoc as a convergent database for the qualitative and quantitative data collected during our formative research and found it to be a comprehensive framework for organizing formative research data. Using the structured constructs of the CFIR allowed us to identify areas that will require additional data collection to ensure future sustainability in the eventual scale-up of the mIVAA system.

Formative research is an iterative process that both informs the constructs of the CFIR domains and serves to strengthen them by formalizing stakeholder input and providing a discussion point for brainstorming solutions. We were able to refer to Tables 1 and 2, created through the use of the CFIR domains and constructs, during the rollout of the subsequent pilot study and ensure that the identified barriers were appropriately addressed. The use of a framework to guide formative research could be a time- and cost-saving measure and provide a systematic mechanism to ensure that facilitators are taken advantage of and barriers are addressed before rollout. However, it is important to recognize that there may be financial implications to addressing barriers before implementation (eg, purchase of additional equipment and updating of software) that should be accounted for by the implementing institution when budgeting for formative research and pilot studies.

In our study, the CFIR framework revealed a conducive implementation climate and readiness for implementation. There was a strong shared understanding of the need for an intervention to reduce loss to follow-up for women after cervical cancer screening as well as significant investment on the part of La Liga in terms of staff, space, and clinic time. In addition, the constructs under the domain of characteristics of individuals were favorable for implementation based on the survey data collected on provider and staff attitudes toward innovation in cervical cancer screening.

We found that the constructs under the domains of intervention characteristics and inner setting were most likely to elicit potential barriers that could be addressed before pilot implementation, which is similar to the findings of the limited previous work using the CFIR in a preimplementation setting [38]. The formal process of formative data collection was the least applicable to the domain of outer setting, which mainly describes the political and organizational environment

surrounding the intervention. Political and organizational information was gathered before our group interviews and observations through conversations with La Liga leadership and desk research. A key need identified through these discussions was clear role definition for the midwives as only trained medical doctors and colposcopists are permitted to review cervical images as per Peruvian regulations. Our findings indicate that prioritizing interview and observation guides that emphasize the constructs within the domains of intervention characteristics and inner setting may maximize the study team's ability to elicit potential barriers and address them before piloting an intervention.

Limitations

This is a case study and may be limited by the study context, but we believe that the process is widely applicable to other work on the design and implementation of mHealth interventions. In addition, our formative work included a relatively small sample size of participants; however, given the overall size of La Liga's organization, this included most stakeholders and representative voices from all involved staff and clinicians. All participants were recruited through La Liga, their employer, including through a study research coordinator who was also an employee of La Liga. Although the participants provided informed consent and were able to decline participation at any time, this may have exerted some influence on their decision to take part in this study. We took multiple precautions to ensure that the study data translated from Spanish were accurate in meaning and tone; however, we acknowledge that there is always a limitation in using translated transcripts as some nuance may be lost.

As a consequence of timing constraints, we were only able to use rapid analysis findings to elicit feedback from participants during the group discussions. We then validated our convergent mixed methods analysis findings with a select few stakeholders from La Liga, including KDMV and JJ. However, we emphasize that, for future work, stakeholder insight is critical, especially when conducting formative research in an LMIC setting. Owing to the post hoc selection of the CFIR as the convergent framework, we only report on implementation factors that emerged in the data. For instance, we did not collect in-depth data on constructs related to the outer setting. Future studies may benefit from structured data collection tools that a priori encompass all CFIR domains.

Conclusions

Formative research can provide useful insights to inform eventual implementation of mHealth interventions. The CFIR framework can be used to map and prioritize potential barriers to the implementation of mHealth interventions revealed during formative work. In our experience with an mHealth-enabled cervical cancer screening device, focusing on formative work exploring constructs under the domains of intervention characteristics and inner setting elicited the most key barriers to implementation. Future mHealth studies may choose to develop data collection tools to specifically query these domains.

Acknowledgments

The authors are grateful to all study participants, members of La Liga Peruana Contra el Cáncer and members of Medical Innovation and Technology for their support of the research presented in this manuscript. They would also like to thank Dr Hayden Bosworth (Duke University Department of Population Health Sciences) for his review of an earlier draft of this manuscript. The research reported in this publication was supported by the Fogarty International Center of the National Institutes of Health under award R21TW010625. HWR and LV were funded by the National Center For Advancing Translational Sciences of the National Institutes of Health under awards TL1TR002555 and KL2TR002554, respectively. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

JJ is a consultant for Merck, which is a manufacturer of human papillomavirus vaccines.

Multimedia Appendix 1
Supplemental tables describing reporting standards, facilitators of implementation, and nonintervention-specific barriers to implementation.
[DOCX File , 37 KB - formative_v6i6e32577_app1.docx]

References

1. Silver L. Smartphone ownership is growing rapidly around the world, but not always equally. Pew Research Center. 2019 Feb 5. URL: https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/ [accessed 2022-06-08]
2. Labrique A, Vasudevan L, Mehl G, Rosskam E, Hyder AA. Digital health and health systems of the future. Glob Health Sci Pract 2018 Oct 10;6(Suppl 1):S1-S4 [FREE Full text] [doi: 10.9745/GHSP-D-18-00342] [Medline: 30305334]
3. Agarwal S, LeFevre AE, Lee J, L'Engle K, Mehl G, Sinha C, WHO mHealth Technical Evidence Review Group. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. BMJ 2016 Mar 17;352:i1174. [doi: 10.1136/bmj.i1174] [Medline: 26988021]

4. Classification of digital health interventions v1.0: a shared language to describe the uses of digital technology for health. World Health Organization. 2018. URL: https://apps.who.int/iris/bitstream/handle/10665/260480/WHO-RHR-18.06-eng.pdf [accessed 2018-04-08]
5. Tomlinson M, Rotheram-Borus MJ, Swartz L, Tsai AC. Scaling up mHealth: where is the evidence? PLoS Med 2013;10(2):e1001382 [FREE Full text] [doi: 10.1371/journal.pmed.1001382] [Medline: 23424286]
6. Lennon MR, Bouamrane MM, Devlin AM, O'Connor S, O'Donnell C, Chetty U, et al. Readiness for delivering digital health at scale: lessons from a longitudinal qualitative evaluation of a national digital health innovation program in the United Kingdom. J Med Internet Res 2017 Feb 16;19(2):e42 [FREE Full text] [doi: 10.2196/jmir.6900] [Medline: 28209558]
7. Schreiweis B, Pobiruchin M, Strotbaum V, Suleder J, Wiesner M, Bergh B. Barriers and facilitators to the implementation of eHealth services: systematic literature analysis. J Med Internet Res 2019 Nov 22;21(11):e14197 [FREE Full text] [doi: 10.2196/14197] [Medline: 31755869]
8. Thornicroft G, Lempp H, Tansella M. The place of implementation science in the translational medicine continuum. Psychol Med 2011 Oct;41(10):2015-2021. [doi: 10.1017/S0033291711000109] [Medline: 21320391]
9. Nilsen P. Making sense of implementation theories, models and frameworks. Implement Sci 2015 Apr 21;10:53 [FREE Full text] [doi: 10.1186/s13012-015-0242-0] [Medline: 25895742]
10. Monitoring and evaluating digital health interventions: a practical guide to conducting research and assessment. World Health Organization. 2016. URL: https://apps.who.int/iris/rest/bitstreams/1069891/retrieve [accessed 2022-06-08]
11. The MAPS Toolkit: mHealth Assessment and Planning for Scale. World Health Organization. 2015. URL: https://apps.who.int/iris/bitstream/handle/10665/185238/9789241509510_eng.pdf [accessed 2022-06-08]
12. Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. Med Care 2012 Mar;50(3):217-226 [FREE Full text] [doi: 10.1097/MLR.0b013e3182408812] [Medline: 22310560]
13. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implement Sci 2016 Oct 26;11(1):146 [FREE Full text] [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
14. Jacob C, Sanchez-Vazquez A, Ivory C. Social, organizational, and technological factors impacting clinicians' adoption of mobile health tools: systematic literature review. JMIR Mhealth Uhealth 2020 Feb 20;8(2):e15935 [FREE Full text] [doi: 10.2196/15935] [Medline: 32130167]
15. Bruni L, Albero G, Serrano B, Mena M, Collado JJ, Gómez D, et al. Human papillomavirus and related diseases in the world. Summary Report. ICO/IARC information centre on HPV and cancer (HPV information centre). 2019. URL: https://hpvcentre.net/statistics/reports/XWX.pdf [accessed 2022-06-08]
16. Arbyn M, Weiderpass E, Bruni L, de Sanjosé S, Saraiya M, Ferlay J, et al. Estimates of incidence and mortality of cervical cancer in 2018: a worldwide analysis. Lancet Glob Health 2020 Feb;8(2):e191-e203 [FREE Full text] [doi: 10.1016/S2214-109X(19)30482-6] [Medline: 31812369]
17. WHO guidelines for screening and treatment of precancerous lesions for cervical cancer prevention. 2nd edition. World Health Organization. 2013. URL: https://apps.who.int/iris/rest/bitstreams/1354685/retrieve [accessed 2022-06-08]
18. Yang D, Soulos PR, Davis B, Gross CP, Yu JB. Impact of widespread cervical cancer screening: number of cancers prevented and changes in race-specific incidence. Am J Clin Oncol 2018 Mar;41(3):289-294 [FREE Full text] [doi: 10.1097/COC.0000000000000264] [Medline: 26808257]
19. Ronco G, Dillner J, Elfström KM, Tunesi S, Snijders PJ, Arbyn M, International HPV screening working group. Efficacy of HPV-based screening for prevention of invasive cervical cancer: follow-up of four European randomised controlled trials. Lancet 2014 Feb 08;383(9916):524-532. [doi: 10.1016/S0140-6736(13)62218-7] [Medline: 24192252]
20. Horn J, Denecke A, Luyten A, Rothe B, Reinecke-Lüthge A, Mikolajczyk R, et al. Reduction of cervical cancer incidence within a primary HPV screening pilot project (WOLPHSCREEN) in Wolfsburg, Germany. Br J Cancer 2019 May;120(10):1015-1022 [FREE Full text] [doi: 10.1038/s41416-019-0453-2] [Medline: 30988395]
21. Mueller JL, Lam CT, Dahl D, Asiedu MN, Krieger MS, Bellido-Fuentes Y, et al. Portable Pocket colposcopy performs comparably to standard-of-care clinical colposcopy using acetic acid and Lugol's iodine as contrast mediators: an investigational study in Peru. BJOG 2018 Sep;125(10):1321-1329 [FREE Full text] [doi: 10.1111/1471-0528.15326] [Medline: 29893472]
22. Lam CT, Krieger MS, Gallagher JE, Asma B, Muasher LC, Schmitt JW, et al. Design of a novel low cost point of care tampon (POCkeT) colposcope for use in resource limited settings. PLoS One 2015 Sep 2;10(9):e0135869 [FREE Full text] [doi: 10.1371/journal.pone.0135869] [Medline: 26332673]
23. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 07;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
24. Kirk MA, Kelley C, Yankey N, Birken SA, Abadie B, Damschroder L. A systematic review of the use of the Consolidated Framework for Implementation Research. Implement Sci 2016 May 17;11:72 [FREE Full text] [doi: 10.1186/s13012-016-0437-z] [Medline: 27189233]

25. Tilahun B, Smillie K, Bardosh KL, Murray M, Fitzgerald M, Cook V, et al. Identifying barriers and facilitators of 13 mHealth projects in North America and Africa: protocol for a 5-year Implementation Science study. JMIR Res Protoc 2018 Jul 03;7(7):e162 [FREE Full text] [doi: 10.2196/resprot.9633] [Medline: 29970360]
26. Meyer AJ, Armstrong-Hough M, Babirye D, Mark D, Turimumahoro P, Ayakaka I, et al. Implementing mHealth interventions in a resource-constrained setting: case study from Uganda. JMIR Mhealth Uhealth 2020 Jul 13;8(7):e19552 [FREE Full text] [doi: 10.2196/19552] [Medline: 32673262]
27. Rodgers M, Thomas S, Harden M, Parker G, Street A, Eastwood A. Developing a Methodological Framework for Organisational Case Studies: A Rapid Review and Consensus Development Process. Southampton, UK: NIHR Journals Library; Jan 2016.
28. Aarons GA. Mental health provider attitudes toward adoption of evidence-based practice: the Evidence-Based Practice Attitude Scale (EBPAS). Ment Health Serv Res 2004 Jun;6(2):61-74 [FREE Full text] [doi: 10.1023/b:mhsr.0000024351.12294.65] [Medline: 15224451]
29. Vasudevan L, Stinnett S, Mizelle C, Melgar K, Makarushka C, Pieters M, et al. Barriers to the uptake of cervical cancer services and attitudes towards adopting new interventions in Peru. Prev Med Rep 2020 Oct 2;20:101212 [FREE Full text] [doi: 10.1016/j.pmedr.2020.101212] [Medline: 33224718]
30. Beebe J. Basic concepts and techniques of rapid appraisal. Human Organ 1995 Mar;54(1):42-51. [doi: 10.17730/humo.54.1.k84tv883mr2756l3]
31. Gale RC, Wu J, Erhardt T, Bounthavong M, Reardon CM, Damschroder LJ, et al. Comparison of rapid vs in-depth qualitative analytic methods from a process evaluation of academic detailing in the Veterans Health Administration. Implement Sci 2019 Feb 01;14(1):11 [FREE Full text] [doi: 10.1186/s13012-019-0853-y] [Medline: 30709368]
32. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
33. Westgard C, Fleming WO. The use of Implementation Science tools to design, implement, and monitor a community-based mHealth intervention for child health in the Amazon. Front Public Health 2020 Aug 19;8:411 [FREE Full text] [doi: 10.3389/fpubh.2020.00411] [Medline: 32974257]
34. Ware P, Ross HJ, Cafazzo JA, Laporte A, Gordon K, Seto E. Evaluating the implementation of a mobile phone-based telemonitoring program: longitudinal study guided by the Consolidated Framework for Implementation Research. JMIR Mhealth Uhealth 2018 Jul 31;6(7):e10768 [FREE Full text] [doi: 10.2196/10768] [Medline: 30064970]
35. Haverhals LM, Sayre G, Helfrich CD, Battaglia C, Aron D, Stevenson LD, et al. E-consult implementation: lessons learned using consolidated framework for implementation research. Am J Manag Care 2015 Dec 01;21(12):e640-e647 [FREE Full text] [Medline: 26760426]
36. Arrossi S, Paolino M, Orellana L, Thouyaret L, Kohler RE, Viswanath K. Mixed-methods approach to evaluate an mHealth intervention to increase adherence to triage of human papillomavirus-positive women who have performed self-collection (the ATICA study): study protocol for a hybrid type I cluster randomized effectiveness-implementation trial. Trials 2019 Feb 26;20(1):148 [FREE Full text] [doi: 10.1186/s13063-019-3229-3] [Medline: 30808379]
37. Kowalski CP, Veeser M, Heisler M. Formative evaluation and adaptation of pre-and early implementation of diabetes shared medical appointments to maximize sustainability and adoption. BMC Fam Pract 2018 Jul 07;19(1):109 [FREE Full text] [doi: 10.1186/s12875-018-0797-3] [Medline: 29981568]
38. Harry ML, Saman DM, Truitt AR, Allen CI, Walton KM, O'Connor PJ, et al. Pre-implementation adaptation of primary care cancer prevention clinical decision support in a predominantly rural healthcare system. BMC Med Inform Decis Mak 2020 Jun 23;20(1):117 [FREE Full text] [doi: 10.1186/s12911-020-01136-8] [Medline: 32576202]

Abbreviations

CFIR: Consolidated Framework for Implementation Research
EBPAS: Evidence-based Practice Attitude Scale
La Liga: La Liga Peruana Contra el Cáncer
LMIC: low- and middle-income country
mHealth: mobile health
mIVAA: mobile Inspección Visual con Ácido Acético
VIA: visual inspection with acetic acid

Edited by A Mavragani; submitted 05.08.21; peer-reviewed by J Lowery, JD Carrier; comments to author 20.10.21; revised version received 03.03.22; accepted 02.05.22; published 23.06.22.

Please cite as:
Reid HW, Proeschold-Bell RJ, Makarushka C, Melgar Vega KD, Huchko M, Jeronimo J, Vasudevan L
Using the Consolidated Framework for Implementation Research to Inform the Design of the Mobile Inspección Visual con Ácido Acético System: Mixed Methods Case Study
JMIR Form Res 2022;6(6):e32577
URL: https://formative.jmir.org/2022/6/e32577
doi:10.2196/32577
PMID:35737455

Original Paper

Women's Preferences and Design Recommendations for a Postpartum Depression Psychoeducation Intervention: User Involvement Study

Shailee Siddhpuria[1], BHSc; Genevieve Breau[2], PhD; Madison E Lackie[1], MSc; Brynn M Lavery[3], BA; Deirdre Ryan[4,5], MB; Barbara Shulman[4,5], MD; Andrea L Kennedy[4,5], MD; Lori A Brotto[3,6], PhD

[1]Department of Undergraduate Medicine, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada
[2]School of Human Sciences, University of Greenwich, London, United Kingdom
[3]Department of Obstetrics & Gynaecology, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada
[4]Reproductive Mental Health Program, British Columbia Women's Hospital and Health Centre, Vancouver, BC, Canada
[5]Department of Psychiatry, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada
[6]Women's Health Research Institute, British Columbia Women's Hospital and Health Centre, Vancouver, BC, Canada

Corresponding Author:
Lori A Brotto, PhD
Department of Obstetrics & Gynaecology
Faculty of Medicine
University of British Columbia
2775 Laurel Street, 6th Floor
Vancouver, BC, V5Z 1M9
Canada
Phone: 1 604 875 4111 ext 68898
Email: lori.brotto@vch.ca

Abstract

Background: Postpartum depression (PPD) is one of the leading causes of maternal morbidity, affecting up to 18% of Canadian new mothers. Yet, PPD often remains untreated due to numerous barriers in access to care, including location and cost. Development of eHealth interventions in collaboration with patient partners offers an exciting opportunity to fill this care gap and provide effective and affordable care to new parents across British Columbia.

Objective: Our aim was to determine the content and design preferences of women previously diagnosed with PPD to inform changes to the development of a web-enabled intervention for education and management of PPD.

Methods: Webpage prototypes were created to mimic the web-enabled resource using findings from completed focus group research that assessed what women want in a web-enabled support resource for PPD. A convenience sample of women aged >18 years and previously diagnosed with PPD was recruited. Feedback was collected on the content and design of the prototypes via semistructured interviews and online surveys. Qualitative, inductive analytic, and quantitative methods were used.

Results: A total of 9 women (mean age 37.2 years, SD 4.8 years) completed the interview and a majority of the survey. The following 6 themes were identified: (1) inefficacy of text-heavy layouts, (2) highlighting key information, (3) clarity/understandability of the language, (4) finding support groups, (5) validation and immediate help for feelings of isolation, and (6) helpfulness and accessibility of the resource. Each theme identified elements of content or design that were either effective or may be improved upon. Most women (8/9, 89%) favored content relating to foundational knowledge of PPD, such as symptoms and management options. The layout, language, and content were found to be generally easy to understand, clear, trustworthy, and helpful.

Conclusions: Six key areas were identified by women previously diagnosed with PPD, as requiring focus in a web-enabled psychoeducation program. Consistent with past research, this study also found that support and enthusiasm for web-enabled programs support PPD management as an adjunct to other evidence-based treatments.

(JMIR Form Res 2022;6(6):e33411) doi:10.2196/33411

KEYWORDS

postpartum; depression; perinatal mental health; patient engagement; women's health; qualitative; psychoeducation; digital tools

Introduction

Background

Perinatal psychiatric disorders are among the leading causes of maternal morbidity and mortality worldwide, with postpartum depression (PPD) symptoms affecting up to 18% of new mothers in Canada [1,2]. Current best practice guidelines recommend management based on the severity of symptoms [3]. In mild to moderate forms of PPD, nonpharmacological treatments are recommended first, which include psychoeducation, self-care, and psychotherapies, such as cognitive behavioral therapy, prior to pharmacological treatments such as antidepressants [3,4]. In particular, psychoeducation, which involves the provision of evidence-based information, is indicated as a first-line option for education, prevention, and treatment purposes for most people experiencing mild to moderate PPD [3,5]. Despite these effective management options, many women are left untreated, often leading to poor maternal and infant health outcomes [6].

Numerous barriers, including social (eg, stigma), instrumental (eg, financial constraint), and structural (eg, lack of accessible information) barriers, prevent mothers experiencing PPD and their partners from receiving appropriate management [7]. Additionally, the COVID-19 pandemic has been shown to elevate the risks of poor mental health symptoms among pregnant and postpartum women [8,9]. With rising mental health concerns and likely further reduced access to care, the needs of individuals experiencing PPD require more attention than ever [10].

Web-enabled interventions (eg, websites) allow for the translation of psychological and other skill-based interventions via a web-based platform. Such interventions can be instrumental in targeting many barriers as they can be more accessible, affordable, and personalized, particularly in perinatal care [11-13]. With additional challenges to access during COVID-19, easily available web-enabled interventions have become even more prevalent [14]. A number of web-enabled interventions for perinatal mental health have been introduced, though with minimal engagement of end users in their development [15-17].

There is evidence that patient partnership in research designed to develop interventions results in an improved end product given that they centralize the users' perspectives and context, as well as their lived experiences [18]. In a recent study, patient education material that was co-created with patients demonstrated a higher usability score and overall preference in comparison to education material created by only experts [19].

As a precursor to this study, Lackie et al [20] conducted focus groups to determine the unmet digital health needs of women with PPD. Participants in the focus groups believed that a web-enabled intervention could address current gaps within PPD care, including education, validation, empowerment, and accessibility [20]. Through this, the next step in this work was identified as the need to create an accessible web-enabled intervention for all women, specifically those in remote communities where in-person resources are limited.

Objectives

In this pilot phase, we aimed to engage end users by evaluating women's feedback on webpage prototypes of PPD-related psychoeducation content that will eventually be housed within the web-enabled intervention. We hypothesize that this prototype will be overall well-received by our participants, with feedback on various aspects including visual design and content that we will employ to improve the final product.

Methods

Study Population and Eligibility

The study participants were a convenience sample of women from any community across British Columbia (BC), Canada, irrespective of city of habitation or ethnicity. The inclusion criteria were established to ensure meaningful involvement within the limitations of the study. To be eligible, the study required participants to be 18 years or older; be able to read, write, and speak conversational English; have a previous diagnosis of PPD in the last 5 years; have no current PPD symptoms; and have access to a computer/device with stable internet connection. Each participant underwent a screening phone call to ensure eligibility, including a researcher-administered Edinburgh Postnatal Depression Scale (score of <12) [21]. Individuals who were eligible, as per an eligibility screening questionnaire, were invited to participate. This was a pilot study to inform the development of a web-based resource, where we aimed to recruit a small sample size for in-depth qualitative analysis.

Recruitment

Participants were a convenience sample of women in the community recruited primarily through the following 2 methods: (1) recruitment of those who had participated in the foundational focus group research project by Lackie et al and provided consent to be contacted for future research [20] and (2) recruitment via social media advertising. We emailed each previous participant who had consented to be contacted again regarding this new phase of the study and had them respond if they wished to learn more or participate. We only reached out to each participant once unless they responded. Social media posts were made in relevant groups that catered to our target population, such as parent groups across BC. These posts provided preliminary information about the study via a study poster and a short blurb including time commitment, and participants contacted the researchers via the email provided in the post if they were interested in participating. All recruitments were completed between July and August 2020.

Procedures

This was a convergent mixed methods design, in which the quantitative and qualitative phases of the study were conducted separately, and the findings were combined in the interpretation

stage, as described by Cresswell and Plano Clark [22]. This method allows for a data validation approach, in which the open-ended questions in the interview validated the closed-ended responses in the survey [22]. To reduce participant fatigue and obtain richer data, these open-ended questions were asked in a video interview and not as open-ended questions with written responses on a survey.

Content Development

PPD-related psychoeducation content was initially co-drafted by a Web Development Advisory Committee (WDAC) that included researchers, psychologists, psychiatrists, community organization representatives, and patient representatives. The content was reviewed by each member thoroughly, and curated to be as relevant as possible to the target users based on recommendations and suggestions made by 2 patient representatives in the WDAC. This group virtually met multiple times to discuss the final content and design of the prototypes. Each member was reimbursed for their contribution. This content was displayed to emulate a web-enabled platform, including added colors and graphics (Multimedia Appendix 1). The pilot material included educational information, provincial resources, and instructions on finding support in the community. The education section housed foundational knowledge of PPD, including symptoms, definitions, and descriptions of management options. The resource section guided users to locate services and relevant online material. The support section had information on how users can support themselves, including creating one's own community group or at-home self-care. The webpage prototypes were created to include only certain elements of each of these sections in order to grasp the breadth of the final website without making the material exhaustive. For example, the support section prototype only included how to create one's own community group. It is important to highlight that these prototypes were displayed to participants in a document format made to emulate the platform, rather than a professional developed version of the web-enabled intervention.

Demographic Questionnaire and Content Survey (Quantitative)

All enrolled participants completed a demographic and content survey virtually from their personal devices by following an individualized link created using Research Electronic Data Capture (REDCap) tools hosted at BC Women's Hospital in BC, Canada [23]. The demographic questionnaire collected personal information, including age, sex, socioeconomic status, relationship status, and medical history. The content survey presented webpage prototypes from each aforementioned section of the drafted content (Multimedia Appendix 2). Participants were asked to rank a list of content topics based on the quantity of content they would like to see from "not much content" to "lots of content." Participants were also asked for their agreement on statements relating to content clarity, novelty, relevance, and usefulness, using a Likert scale from "strongly disagree" to "strongly agree." Three binary (yes/no) questions were asked regarding the effectiveness of the visual aspects. Overall acceptability was ranked on a Likert scale from "very dissatisfied" to "very satisfied."

Quantitative analysis was conducted with the Statistical Package for the Social Sciences, and descriptive statistics, including mean, standard deviation, and frequency counts and percentages, were calculated [24].

Videoconferencing Interview (Qualitative)

Interviews were virtual and semistructured, consisting of broad open-ended questions and participant-guided discussions on the webpage prototypes, and were carried out by 2 members of the research team, with 1 as the primary facilitator. The principal investigator, who is a registered psychologist, was on-call during all interviews to mediate any high-risk situations if needed. Interviews were conducted with 1 participant at a time to ensure anonymity and comfort. Each participant was given a briefing on the anonymity of the interview and was encouraged to share thoughts openly and without fear of repercussion. All participants were asked to refrain from using any personally identifying information, such as names or geographical locations, during the interview session.

An interview guide was created by the research team. The interview was initiated with an open-ended question regarding their overall feelings about the content. Each participant was also asked to compare the content with previous expectations and infer relevance and usefulness based on their lived experiences. Interviews were concluded with an open question about any final thoughts. Participants were free to discuss any thoughts relating to the intervention at any point during the interview. All interviews were audio recorded and transcribed by a professional transcriptionist (Multimedia Appendix 3).

Qualitative inductive analysis was conducted following the recommendations by Thorne [25]. In vivo coding was conducted on initial interview transcripts, whereby categories were created based on key phrases or words used frequently by participants. These categories were then combined until themes emerged from the data. The coder was a student who trained on qualitative analysis for this project, worked within a large research team at the Women's Health Research Institute in BC, Canada, and verified the data or resolved any uncertainties. The coder was supervised by a PhD-trained researcher who had specialized training in qualitative methodology. The first author (SS) also maintained rigorous research through writing memos regarding the coding throughout the analysis process and consulted with the PhD researcher throughout the process. The interview data were also triangulated with data from the survey phase of the study. Any differences were resolved through consensus.

Ethics Approval

Ethics approval (#H20-00931) was obtained from the University of British Columbia and Children's and Women's Research Ethics Board. Written informed consent was obtained from all participants prior to participating. The confidentiality of the participants was maintained at all times.

Results

Demographics

A total of 9 women consented to participate in the semistructured interview and questionnaire survey (Table 1). The mean age of the participants was 37 years, with an average of 2 children per participant. A majority of the participants were White European, while a minority of the participants included those with Chinese, Indigenous, and South Asian backgrounds. All participants were BC residents, married, heterosexual, and well-educated. A majority of the participants had an average household income of CAD $100,000 or more (US $78,700 or more). One participant did not complete the entire survey; thus, the analyses pertaining to the support section of the webpage prototypes were available only for 8 participants.

Table 1. Sociodemographic characteristics of the participants (N=9).

Variable	Value[a]
Age (years), mean (SD)	37.2 (4.76)
Sex assigned at birth, n (%)	
Female	9 (100)
Gender identity, n (%)	
Woman	9 (100)
Sexual orientation, n (%)	
Heterosexual	9 (100)
Ethnicity, n (%)[b]	
Chinese	1 (11)
Indigenous	1 (11)
South Asian (East Indian, Pakistani, Sri Lankan, etc)	1 (11)
White European	7 (78)
Education, n (%)	
Attended some college/university	1 (11)
Graduated 4-year college/university	4 (44)
Postgraduate degree	4 (44)
Years spent at school or in full-time study, mean (SD)	16.7 (2.69)
Employment, n (%)[b]	
Full time	5 (56)
Part time	2 (22)
On maternity leave	2 (22)
Self-employed	1 (22)
Annual household income, n (%)	
CAD $99,999 or less (US $78,699 or less)	3 (33)
CAD $100,000 or more (US $78,700 or more)	6 (56)
Relationship status, n (%)	
Married	9 (100)
Average length of current relationship (years), mean (SD)	12.6 (6.29)
Average length of longest relationship (months), mean (SD)	12.6 (6.29)
Number of children, mean (SD)	1.8 (0.44)

[a]Percentages may not equal 100 due to rounding.

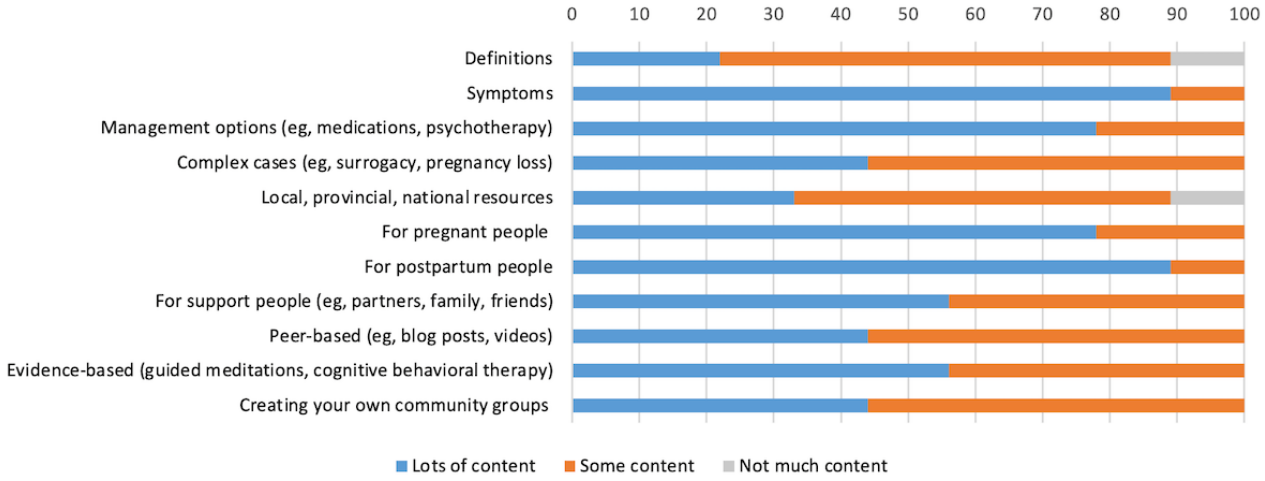
[b]These data include multiple responses from individual participants.

Survey (Quantitative) Results

Overall Content Preferences

Nearly all participants (8/9, 89%) preferred to see "lots of content" relating to symptoms of PPD and content tailored for postpartum people (Figure 1). A majority of the participants (7/9, 78%) preferred to see "lots of content" relating to management options and tailored for pregnant people. Definitions relating to PPD and information on local, provincial, and national resources were least preferred among the participants (3/9, 33%), who preferred either "not much content" or "only some content." Other categories were more evenly distributed between "some content" and "lots of content," such as information around peer-based supports (4/9, 44%).

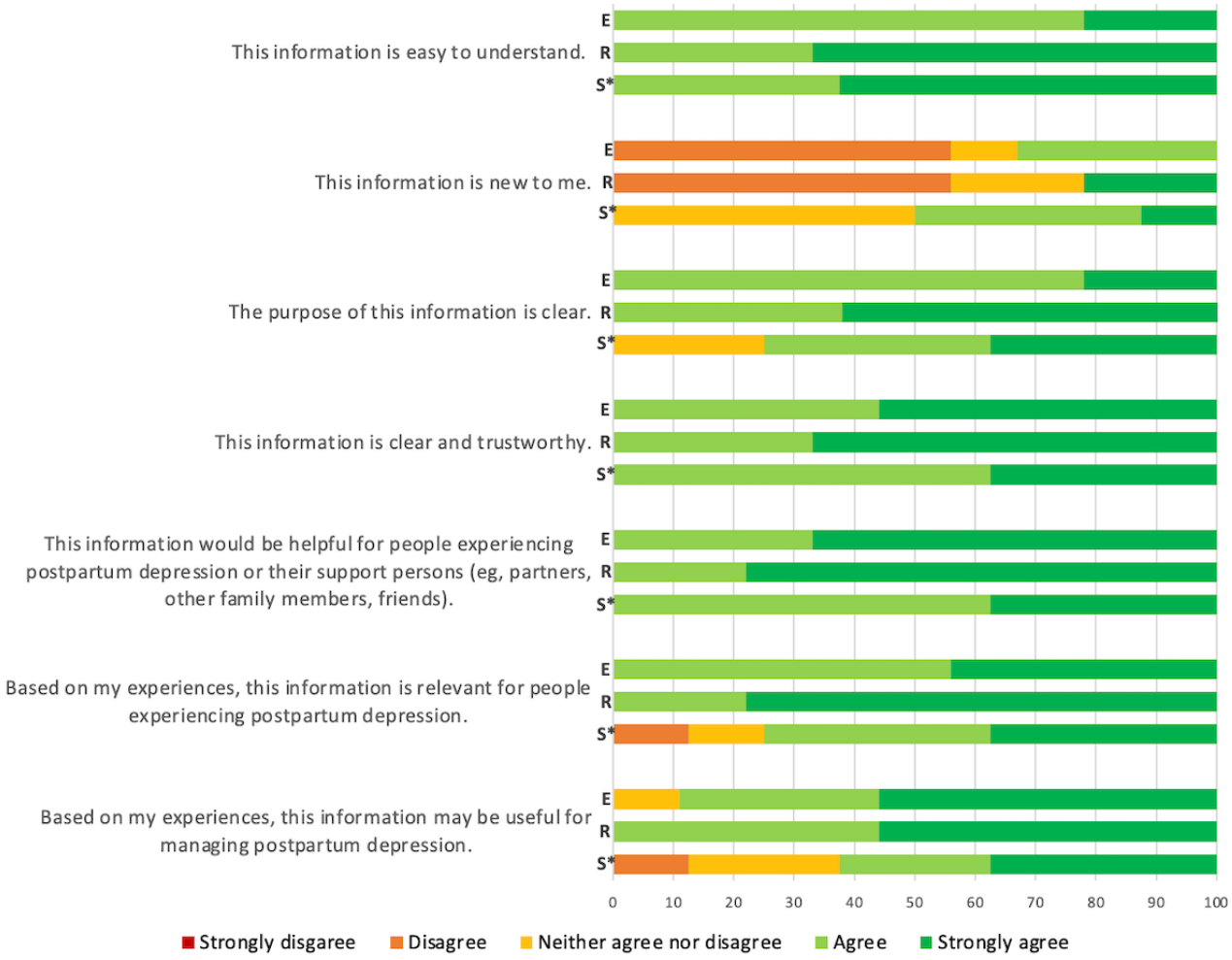
Figure 1. Participant preferred content ratings from "not much content" to "lots of content".



Section-Specific Content Feedback

Nearly all participants (8-9/9, 89%-100%) either agreed or strongly agreed that the information in each webpage prototype section was easy to understand, clear and trustworthy, and helpful to people experiencing PPD and their supports (Figure 2). The information in each section was not new (ie, had pre-existing knowledge) for approximately 50% or more of the participants. All participants (9/9, 100%) agreed or strongly agreed that the purposes of the education and resource webpage prototypes were clear. A majority of the participants (6/8, 75%) agreed or strongly agreed that the purpose of the support excerpt was clear, while a few (2/8, 25%) neither agreed nor disagreed. Finally, nearly all participants (8-9/9, 89%-100%) indicated that the information in the education and resource webpage prototypes would be relevant for people experiencing PPD and helpful for its management. A majority of the participants (5-6/9, 55%-67%) found that the information in the support excerpt was relevant to people experiencing PPD and helpful for its management, while 1 participant (1/8, 13%) disagreed and a few participants (1-2/8, 13%-25%) neither agreed nor disagreed.

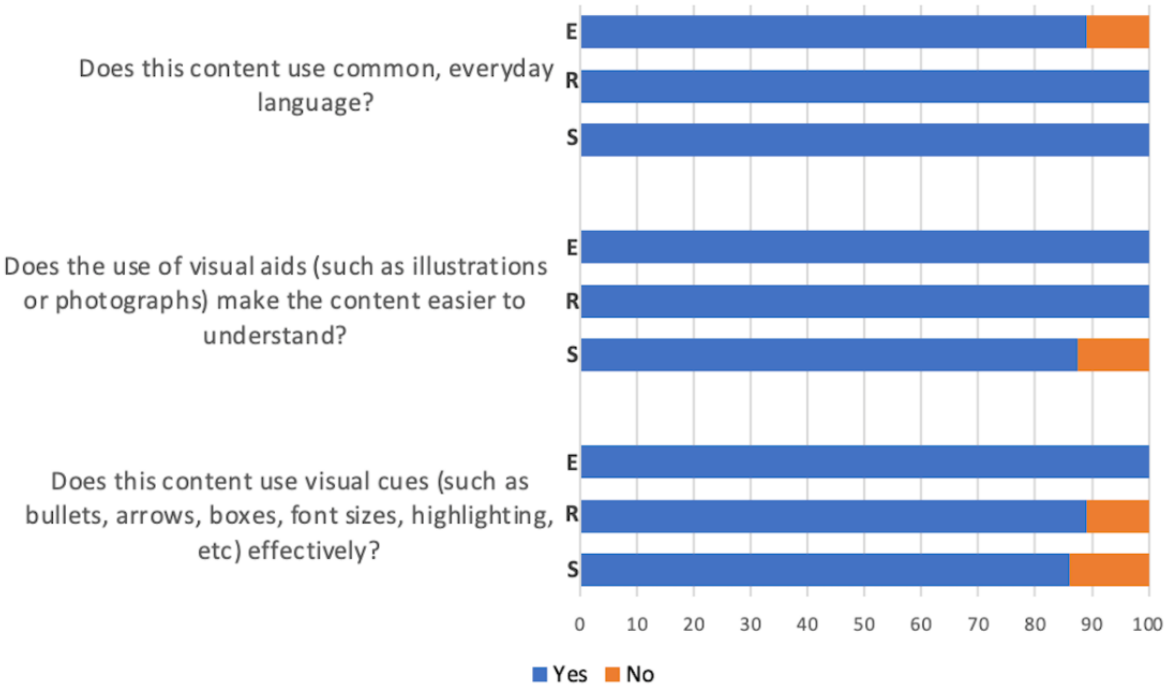
Figure 2. Survey responses to the excerpt content by section using a Likert scale (N=9). For the support webpage prototypes, N=8 as 1 participant did not complete this section. E: education excerpt; R: resource webpage prototypes; S: support webpage prototypes.



Section-Specific Language/Layout Feedback

All participants (9/9 or 8/8, 100%) found that common everyday language was used in the resources and support sections, while most (8/9, 89%) found that to be true for the education section (Figure 3). All participants (9/9, 100%) agreed that visual aids made the content easier to understand in the education and resources sections, while most (7/8, 88%) found that to be true for the support section. All participants (9/9, 100%) agreed that visual cues, such as font sizes, bullets, and bold style, were effectively used in the education and resources sections, while most (7/8, 88%) found that to be true for the support section.

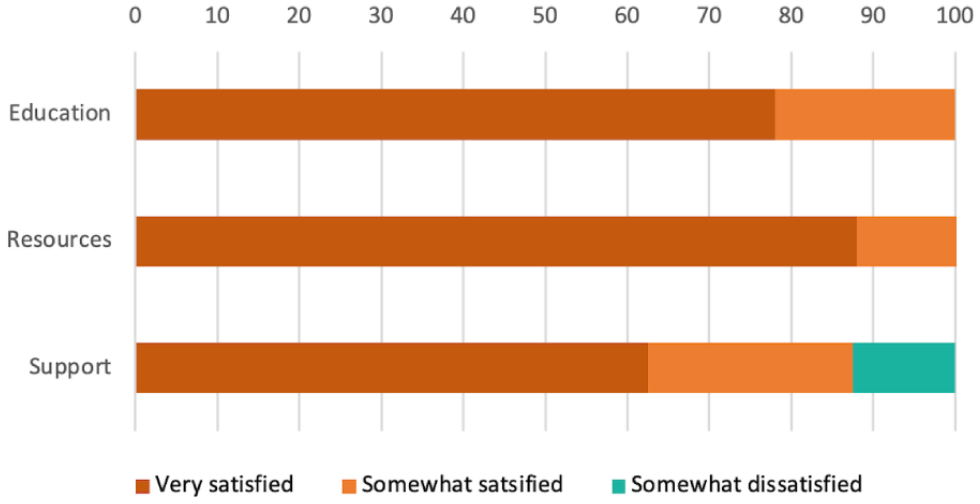
Figure 3. Survey responses to the excerpt language and layout style by section. E: education excerpt; R: resource webpage prototypes; S: support webpage prototypes.



Overall Satisfaction

All participants were either somewhat satisfied (1-2/9, 11%-22%) or very satisfied (7-8/9, 78%-89%) with the content and presentation of the education and resources sections (Figure 4). While most participants were either somewhat satisfied (2/8, 25%) or very satisfied (5/8, 63%) with the content and presentation of the support section, 1 participant (1/8, 13%) was somewhat dissatisfied.

Figure 4. Overall satisfaction with each section excerpt.



Open-Ended Interview (Qualitative) Results: Themes

Each videoconference interview ranged from 20 to 30 minutes. Using inductive analysis, we identified the following 6 themes [25]: (1) highlighting key information, (2) inefficacy of text-heavy layouts, (3) clarity/understandability of the language, (4) finding support groups, (5) validation and immediate help for feelings of isolation, and (6) helpfulness and accessibility of the resource. Both the strengths of the webpage prototypes and current gaps were discussed within each theme.

Highlighting Key Information

Generally, participants agreed that drawing attention to key information is pertinent for future users experiencing PPD. Participants frequently identified the need to be able to "grab something useful quickly" by using stylistic techniques such as "big bold letters," "certain lines being highlighted," "bullet points," and "buttons." Multiple respondents expressed being drawn to attention grabbing cues, and felt that this was important, especially when feeling overwhelmed.

> *I would increase the font…when you have depression and anxiety, it's just very hard to focus on certain*

> *things…key words are almost better…bolding certain things maybe.* [Participant aged 38 years]
>
> *Really great use of bullet points, it wasn't overwhelming, you could scan through it and kind of see what jumps out at you.* [Participant aged 44 years]
>
> *I like how it's clear, and if I want to access something really quickly, there's big bold letters and pictures.* [Participant aged 37 years]

Thus, while the webpage prototypes had some attention-grabbing elements, participants noted specific techniques that can be implemented throughout the resource or enhanced further on the displayed pages.

Inefficacy of Text and Information-Dense Layout

Most participants said that text and information-dense pages are ineffective, while some indicated that having all of the information available is also important. Given the identified need of information that draws attention, text-heavy pages contrast this need by diffusing attention to large areas of text. Multiple participants indicated that on some pages, "there's a lot of text" and "it's a bit too wordy," and that it may be overwhelming, especially for people experiencing PPD.

> *I would want it to be very user friendly, quickly just dive into questions that you need answered, and not a lot of red tape to get through. Often when you are in the throes of postpartum, red tape is not helpful…typically don't have the energy for it.* [Participant aged 38 years]
>
> *I felt like there was a lot of information there, but I actually thought it was kind of overwhelming…long paragraph, small font, it's hard to read. I'd probably wouldn't read it.* [Participant aged 36 years]
>
> *It's a bit too wordy…from a marketing perspective, people only read like the first 10 seconds of something and then move on.* [Participant aged 38 years]

Generally, it was agreed that "hiding the text" and giving users the choice to view it can achieve a balance. Thus, information-dense pages can be overwhelming for users, but given the importance of the information, strategies can be used to reduce overall text.

Clarity and Understandability of the Language

Half of the participants found the language throughout clear and easy to follow, while the other half raised concerns that the language may be too technical at times and thus inaccessible for some individuals. Generally, participants agreed that common easy words should be used throughout the resource as it would increase the accessibility of the resource and simplify content for when amidst PPD.

> *I really liked how you used a common word and then also have the health or medical word for it, because for someone like me, when I go to my healthcare provider, I feel like I can understand this. You're bridging the gap of understanding for me.* [Participant aged 38 years]
>
> *I like the language and visuals. They are not medical, like difficult to read. They're just simple. And when you are in postpartum depression, you kind of need things simplified, right?* [Participant aged 37 years]
>
> *The language may not be easily understandable by the common population because it's got some really big words…it's very scientific.* [Participant aged 36 years]
>
> *In some ways, [the language] is quite technical…some things need to be, with medication, it needs to be in formal language but maybe others don't need to be so formal.* [Participant aged 36 years]

Another participant acknowledged that English being one's first language may limit the ability to determine the clarity of the language for users who speak English as a second or more language.

> *With English as my first language…I thought [the language] was quite straightforward. I found any time you had short forms or abbreviations, I could find it right away.* [Participant aged 38 years]

Thus, participants agreed that language should be accessible to all individuals, but complex language may be incorporated alongside it.

Finding Support Groups

Generally, participants stated that information on how to find existing support groups should be highlighted in comparison to how to start a support group, which may be less relevant to individuals currently experiencing PPD.

> *Community support group piece is probably not that relevant to me, just because I know I wouldn't do that.* [Participant aged 38 years]
>
> *I feel like if somebody has PPD, it's a lot for them to start to run a support group. For me, what would have been more helpful is being able to link with community support groups…you wanna make an equal amount or more space to finding a support group.* [Participant aged 44 years]
>
> *I didn't particularly think running a support community group was at the forefront of most people going through PPD's mind.* [Participant aged 36 years]

Some also noted the challenges and nuances of starting community groups.

> *It can feel really overwhelming to try to start your own…The idea of starting your own group feels really inaccessible.* [Participant aged 44 years]
>
> *To me, [the content] kind of seemed to gloss over the importance of needing qualified people to run a group like this…Having just that trained facilitator who knew how to allow people to talk in a healthy way, I think that made a huge difference.* [Participant aged 38 years]

Thus, participants found starting a community group lower in priority and feasibility, and suggested instead to include more information regarding finding a support group.

Validation and Immediate Help for Feelings of Isolation

Most participants indicated that feelings of isolation should be further addressed by sharing others' experiences or providing immediate contacts. Overwhelmingly, participants emphasized the need for connection and validation for individuals experiencing PPD, for example, through "blog stories" or by adding a "more humane angle to [the content]" in order to remain hopeful.

> *Just hearing other people articulate their mental and emotional state…was so helpful…so you don't feel so isolated. Having that basis of language of how to talk about…this is what's happening in your brain, on a biophysical level.* [Participant aged 38 years]
>
> *Something that was really helpful for me was connecting with other people that had postpartum depression…who can tell you there's hope at the other end.* [Participant aged 27 years]
>
> *I'd like to see some testimonies, some real-life people sharing stories…that's really comforting to people as well…makes it more human and less clinical.* [Participant aged 36 years]
>
> *You want something that helps them feel like there's some positivity at the end of this, like there's a light at the end of the tunnel.* [Participant aged 44 years]

Many also noted having easy-to-find contact information for the immediate need of an individual experiencing significant distress due to PPD.

> *Usually you reach out for help when things are really bad…and you just need kind of an immediate thing to calm, bring that anxiety down.* [Participant aged 38 years]
>
> *Suicide prevention [phone] line on the top, on every page…having that resource handy and knowing someone will be there to catch you if you fall so you don't really have to fall completely.* [Participant aged 36 years]

Thus, feelings of isolation were commonly experienced by participants, and it was noted that it may be helpful to incorporate relatable stories and validation messages throughout. Moreover, providing contacts for immediate need can help for when isolation feels unbearable.

Helpfulness and Accessibility of the Intervention

Generally, participants believed that the proposed web-enabled intervention would be a helpful and accessible resource to manage PPD initially or in early stages. Many indicated that the intervention would have been "relevant and useful" when they were managing their own PPD and would recommend it to others given the content shown.

> *In my own experience, I had a very difficult time accessing help, and something like this would've been really incredibly helpful for that. I think the fact that it's kind of rarely available…and comes from a well-known place, like it's here. It's BC. It's Canada, there's a lot of weight of that behind it. It's really accessible and…makes you feel not so alone in it, I think that's the impression I was left with.* [Participant aged 44 years]
>
> *I wish I could just hand [the website] out to so many people I know, so when this is up and running…I love that I have something that I can refer to for people I know.* [Participant aged 38 years]
>
> *Useful…especially for people who have never experienced it, and don't know how to or where to turn, or if they feel that they don't want to reach out.* [Participant aged 33 years]

Thus, overall, the content shown was well received by the participants, who noted its potential usefulness not only in their past journeys, but also for others.

Discussion

Principal Findings

This study explored the preferences and recommendations of 9 women previously diagnosed with PPD for a web-enabled psychoeducational intervention. Based on their experiences, all participants generally believed that the presented webpage prototypes would be relevant and useful as an additional resource for education and early management of PPD. Several suggestions emerged as themes, including highlighting key information and increasing the succinctness and clarity of the language, as well as focusing the content on finding help and validation to increase the relevance, accessibility, and user friendliness of the presented prototype. This is in keeping with the findings of an earlier study, where Lackie et al determined that a psychoeducational tool could address unmet gaps in current PPD care [20]. Unlike many existing interventions, the current one is built upon patient engagement during each phase of its development [15,16,26-29].

The quality assessment of web-based tools can be divided into the following 2 major aspects: content and design [30]. Results of this study will be discussed using those 2 aspects and the factors within.

Participants expressed the most preference for the content, and in particular, information relating directly to the symptoms (ie, presentation) and management options for PPD. Several studies have identified a lack of foundational knowledge, such as recognition of symptoms, as a common and significant barrier to seeking necessary help [7,31]. In fact, directing such information to *all* individuals at risk may not just lead to early recognition but even be preventive [32-34]. In contrast, information regarding creating one's own support group may be less relevant. The participants believed that addressing the need for immediate help, such as by helping to locate an *existing* support group, should be prioritized. In the case of PPD, when symptoms and access to care are often already prohibitive, it is important to highlight information that is likely to provide assistance in the most efficient way [31,35]. Furthermore, many participants described feeling alone during their PPD experiences and strongly suggested incorporating validating statements and relatable stories wherever possible. Even amidst social support, feelings of isolation often prevail in PPD [36]. Receiving validation and assurance, particularly from sources

with an understanding of PPD, has been shown to further facilitate help-seeking behavior [37]. Validating messaging and peer stories of lived experiences are just some ways to address this need.

In regard to design, participants in this study unequivocally preferred a simple and easy-to-navigate layout with helpful features. Some specific suggestions included highlighting key information using typography techniques or having the option to view information as needed, rather than all at once. This is congruent with digital health information guidelines, which highlight the importance of enhancing accessibility through design considerations [38]. To cater to a wide range of health literacy skills (referring to an individual's ability to locate, understand, and apply information) between users (ie, anyone with PPD or at risk for PPD), interventions must facilitate easy location and avoid high text density layouts [38,39]. As brought forth by the participants, the possible overwhelming and distressing feelings faced during PPD make it even more important to consider accessibility as the utmost priority in design across this and all perinatal mental health interventions.

Users as Collaborators

Patient-oriented research emphasizes the importance of patient engagement to ensure high relevance, acceptance, and impact of the undertaken project [40]. This study was driven by continued involvement of relevant stakeholders, such as the WDAC (consisting of researchers, psychologists, psychiatrists, community organization representatives, and patients), to elucidate user perspective and context, ultimately informing the implementation of key ideas directly into the intervention. Particularly, each participant in this study had lived experience of PPD, thus making for a highly relevant patient group. Few of the published digital PPD resource studies have incorporated meaningful patient engagement during their development process to inform iterative changes to this degree [17,26,41]. Similar methods as described in this phase have been employed outside of PPD or in the postdevelopment phase for improvement purposes [42,43]. Yet, research has clearly identified enhanced perceived usability and consequently longer duration of use through early and persistent user involvement in development [44]. With increasing reliance of web- and mobile-enabled interventions in many domains of health, it is important to make patient engagement commonplace, as it is often both feasible and highly beneficial.

Currently Available Resources

To date, several web- and mobile-enabled psychoeducation resources can be easily found through search engines, for perinatal mental health or PPD alone. However, few of these have been evaluated in the literature for quality assurance purposes. Of those evaluated, the quality of information still remains low, with several sources not disclosing key elements that help validate the content, such as the sponsorship or authorship [45,46]. Consistent with a previous study that utilized focus groups [20], here we found support again for well-researched and patient-centered digital resources for PPD.

Limitations

While the findings from this study will improve the development of the web-based tool, several limitations of the study must be noted. First, a small sample was recruited due to the constraints of time and method of recruitment, which was primarily convenience sampling. Moreover, the participants were predominantly white, well-educated, and heterosexual married women. This lack of representative sampling limits the overall generalizability of the results. Additionally, participants were shown webpage prototypes in a document format created by the study team rather than a professionally developed version of the web-enabled intervention, which did not allow us to assess participant experiences with navigation. Access to the internet and internet-compatible devices was required for eligibility as we were unable to provide alternative methods of participation owing to the COVID-19 pandemic. This may have disproportionately excluded some individuals. Lastly, this study was prone to the standard limitations of online surveys and interviewing in research, including validity and reliability issues [47,48].

Future Directions

This study adds to the existing literature calling for more evidence-based web-enabled resources for PPD. A resource, such as ours, provides information about PPD and its treatment, links users to community resources, and provides information about recovery from PPD. The next steps for our research team are to conduct a study to evaluate the effectiveness of the web-enabled resource for alleviating the symptoms of PPD in a demographically diverse group of users. Eventually, it is hoped that this will be a universally accessible resource for anyone with access to the internet and a device.

Conclusion

This study reports many specific end-user preferences of women previously diagnosed with PPD to directly inform changes to a web-enabled psychoeducation intervention. Participants generally commended on the perceived helpfulness, reliability, and user friendliness of the webpage prototypes' content and design. They also provided thoughtful suggestions that may enhance the impact and user experience of this resource. This study also further demonstrates the methodology and importance of involving patients at each phase of health intervention development. Overall, there is continued support and hopefulness for the potential role of this intervention as an addition to existing professional care options.

Acknowledgments

We would like to especially thank our funders the Gillespie Family Foundation, the Diamond Foundation, and the British Columbia Women's Health Foundation for funding the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Webpage prototypes as shown to participants.
[PDF File (Adobe PDF File), 2922 KB - formative_v6i6e33411_app1.pdf]

Multimedia Appendix 2
Participant survey based on webpage prototypes.
[PDF File (Adobe PDF File), 139 KB - formative_v6i6e33411_app2.pdf]

Multimedia Appendix 3
Videoconference interview guide.
[PDF File (Adobe PDF File), 100 KB - formative_v6i6e33411_app3.pdf]

References

1. Oates M. Perinatal psychiatric disorders: a leading cause of maternal morbidity and mortality. Br Med Bull 2003;67(1):219-229. [doi: 10.1093/bmb/ldg011] [Medline: 14711766]
2. Maternal mental health in Canada. Statistics Canada. 2019. URL: https://publications.gc.ca/collections/collection_2019/statcan/11-627-m/11-627-m2019041-eng.pdf [accessed 2022-02-01]
3. Williams J. Best Practice Guidelines for Mental Health Disorders in the Perinatal Period. Reproductive Mental Health Program & Perinatal Services. 2014. URL: https://reproductivementalhealth.ca/sites/default/files/uploads/resources/files/mental_health_guidline_web.pdf [accessed 2022-02-01]
4. MacQueen GM, Frey BN, Ismail Z, Jaworska N, Steiner M, Lieshout RJV, CANMAT Depression Work Group. Canadian Network for Mood and Anxiety Treatments (CANMAT) 2016 clinical guidelines for the management of adults with major depressive disorder: Section 6. Special populations: Youth, women, and the elderly. Can J Psychiatry 2016 Sep 11;61(9):588-603 [FREE Full text] [doi: 10.1177/0706743716659276] [Medline: 27486149]
5. Donker T, Griffiths KM, Cuijpers P, Christensen H. Psychoeducation for depression, anxiety and psychological distress: a meta-analysis. BMC Med 2009 Dec 16;7(1):79 [FREE Full text] [doi: 10.1186/1741-7015-7-79] [Medline: 20015347]
6. Slomian J, Honvo G, Emonts P, Reginster J, Bruyère O. Consequences of maternal postpartum depression: A systematic review of maternal and infant outcomes. Womens Health (Lond) 2019 Apr 29;15:1745506519844044 [FREE Full text] [doi: 10.1177/1745506519844044] [Medline: 31035856]
7. Jones A. Help seeking in the perinatal period: A review of barriers and facilitators. Soc Work Public Health 2019 Jun 26;34(7):596-605. [doi: 10.1080/19371918.2019.1635947] [Medline: 31242074]
8. Hessami K, Romanelli C, Chiurazzi M, Cozzolino M. COVID-19 pandemic and maternal mental health: a systematic review and meta-analysis. J Matern Fetal Neonatal Med 2020 Dec 01:1-8. [doi: 10.1080/14767058.2020.1843155] [Medline: 33135523]
9. Davenport MH, Meyer S, Meah VL, Strynadka MC, Khurana R. Moms are not OK: COVID-19 and maternal mental health. Front Glob Womens Health 2020 Jun 19;1:1 [FREE Full text] [doi: 10.3389/fgwh.2020.00001] [Medline: 34816146]
10. Vigod SN, Brown HK, Huang A, Fung K, Barker LC, Hussain-Shamsy N, et al. Postpartum mental illness during the COVID-19 pandemic: a population-based, repeated cross-sectional study. CMAJ 2021 Jun 07;193(23):E835-E843 [FREE Full text] [doi: 10.1503/cmaj.210151] [Medline: 34099467]
11. Eysenbach G. What is e-health? J Med Internet Res 2001 Jun 18;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
12. van den Heuvel JF, Groenhof TK, Veerbeek JH, van Solinge WW, Lely AT, Franx A, et al. eHealth as the next-generation perinatal care: An overview of the literature. J Med Internet Res 2018 Jun 05;20(6):e202 [FREE Full text] [doi: 10.2196/jmir.9262] [Medline: 29871855]
13. Jiao N, Zhu L, Chong YS, Chan WS, Luo N, Wang W, et al. Web-based versus home-based postnatal psychoeducational interventions for first-time mothers: A randomised controlled trial. Int J Nurs Stud 2019 Dec;99:103385. [doi: 10.1016/j.ijnurstu.2019.07.002] [Medline: 31442783]
14. Bierbooms JJPA, van Haaren M, IJsselsteijn WA, de Kort YAW, Feijt M, Bongers IMB. Integration of online treatment into the "new normal" in mental health care in post-COVID-19 times: Exploratory qualitative study. JMIR Form Res 2020 Oct 08;4(10):e21344 [FREE Full text] [doi: 10.2196/21344] [Medline: 33001835]
15. O'Mahen HA, Richards DA, Woodford J, Wilkinson E, McGinley J, Taylor RS, et al. Netmums: a phase II randomized controlled trial of a guided Internet behavioural activation treatment for postpartum depression. Psychol. Med 2013 Oct 09;44(8):1675-1689. [doi: 10.1017/s0033291713002092]

16. Milgrom J, Danaher BG, Gemmill AW, Holt C, Holt CJ, Seeley JR, et al. Internet cognitive behavioral therapy for women with postnatal depression: A randomized controlled trial of MumMoodBooster. J Med Internet Res 2016 Mar 07;18(3):e54 [FREE Full text] [doi: 10.2196/jmir.4993] [Medline: 26952645]
17. Haga SM, Drozd F, Brendryen H, Slinning K. Mamma mia: a feasibility study of a web-based intervention to reduce the risk of postpartum depression and enhance subjective well-being. JMIR Res Protoc 2013 Aug 12;2(2):e29 [FREE Full text] [doi: 10.2196/resprot.2659] [Medline: 23939459]
18. van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011 Dec 05;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
19. Bashir N, Moore J, Buckland D, Rodrigues M, Tonelli M, Thombs B, et al. Are patient education materials about cancer screening more effective when co-created with patients? A qualitative interview study and randomized controlled trial. Curr Oncol 2019 Apr 01;26(2):124-136 [FREE Full text] [doi: 10.3747/co.26.4621] [Medline: 31043815]
20. Lackie ME, Parrilla JS, Lavery BM, Kennedy AL, Ryan D, Shulman B, et al. Digital health needs of women with postpartum depression: Focus group study. J Med Internet Res 2021 Jan 06;23(1):e18934 [FREE Full text] [doi: 10.2196/18934] [Medline: 33404506]
21. Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale. Br J Psychiatry 1987 Jul 02;150(6):782-786. [doi: 10.1192/bjp.150.6.782] [Medline: 3651732]
22. Cresswell JW, Clark VLP. Designing and Conducting Mixed Methods Research. Thousand Oaks, CA: SAGE Publications, Inc; 2011.
23. REDCap. URL: https://projectredcap.org/software/ [accessed 2022-02-01]
24. IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp; 2017.
25. Thorne S. Interpretive Description: Qualitative Research for Applied Practice. Walnut Creek, CA: Left Coast Press; 2008.
26. Danaher BG, Milgrom J, Seeley JR, Stuart S, Schembri C, Tyler MS, et al. MomMoodBooster web-based intervention for postpartum depression: feasibility trial results. J Med Internet Res 2013 Dec 04;15(11):e242 [FREE Full text] [doi: 10.2196/jmir.2876] [Medline: 24191345]
27. Forsell E, Bendix M, Holländare F, Szymanska von Schultz B, Nasiell J, Blomdahl-Wetterholm M, et al. Internet delivered cognitive behavior therapy for antenatal depression: A randomised controlled trial. J Affect Disord 2017 Oct 15;221:56-64 [FREE Full text] [doi: 10.1016/j.jad.2017.06.013] [Medline: 28628768]
28. Lau Y, Cheng J, Wong S, Yen K, Cheng L. Effectiveness of digital psychotherapeutic intervention among perinatal women: A systematic review and meta-analysis of randomized controlled trials. World J Psychiatry 2021 May 19;11(4):133-152 [FREE Full text] [doi: 10.5498/wjp.v11.i4.133] [Medline: 33889538]
29. Mu T, Li Y, Xu R, Chen J, Wang Y, Shen C. Internet-based interventions for postpartum depression: A systematic review and meta-analysis. Nurs Open 2021 May 29;8(3):1125-1134 [FREE Full text] [doi: 10.1002/nop2.724] [Medline: 33373101]
30. Aladwani AM, Palvia PC. Developing and validating an instrument for measuring user-perceived web quality. Information & Management 2002 May;39(6):467-476. [doi: 10.1016/s0378-7206(01)00113-6]
31. Button S, Thornton A, Lee S, Shakespeare J, Ayers S. Seeking help for perinatal psychological distress: a meta-synthesis of women's experiences. Br J Gen Pract 2017 Aug 28;67(663):e692-e699. [doi: 10.3399/bjgp17x692549]
32. Rowe HJ, Fisher JR. Development of a universal psycho-educational intervention to prevent common postpartum mental disorders in primiparous women: a multiple method approach. BMC Public Health 2010 Aug 18;10(1):499 [FREE Full text] [doi: 10.1186/1471-2458-10-499] [Medline: 20718991]
33. Moshki M, Baloochi Beydokhti T, Cheravi K. The effect of educational intervention on prevention of postpartum depression: an application of health locus of control. J Clin Nurs 2014 Aug 12;23(15-16):2256-2263. [doi: 10.1111/jocn.12505] [Medline: 24329943]
34. Lara MA, Navarro C, Navarrete L. Outcome results of a psycho-educational intervention in pregnancy to prevent PPD: a randomized control trial. J Affect Disord 2010 May;122(1-2):109-117. [doi: 10.1016/j.jad.2009.06.024] [Medline: 19596446]
35. Posmontier B. Functional status outcomes in mothers with and without postpartum depression. J Midwifery Womens Health 2008;53(4):310-318 [FREE Full text] [doi: 10.1016/j.jmwh.2008.02.016] [Medline: 18586183]
36. Thomas LJ, Scharp KM, Paxman CG. Stories of postpartum depression: exploring health constructs and help-seeking in mothers' talk. Women Health 2014 May 16;54(4):373-387. [doi: 10.1080/03630242.2014.896442] [Medline: 24617862]
37. Evans M, Donelle L, Hume-Loveland L. Social support and online postpartum depression discussion groups: a content analysis. Patient Educ Couns 2012 Jul;87(3):405-410. [doi: 10.1016/j.pec.2011.09.011] [Medline: 22019021]
38. Health Literacy Online. URL: https://health.gov/healthliteracyonline/ [accessed 2022-02-01]
39. Beaunoyer E, Arsenault M, Lomanowska AM, Guitton MJ. Understanding online health information: Evaluation, tools, and strategies. Patient Educ Couns 2017 Feb;100(2):183-189. [doi: 10.1016/j.pec.2016.08.028] [Medline: 27595436]
40. Strategy for Patient-Oriented Research. Canadian Institutes of Health Research. URL: https://cihr-irsc.gc.ca/e/41204.html [accessed 2022-02-01]
41. Fealy S, Chan S, Wynne O, Dowse E, Ebert L, Ho R, et al. The Support for New Mums Project: A protocol for a pilot randomized controlled trial designed to test a postnatal psychoeducation smartphone application. J Adv Nurs 2019 Jul 21;75(6):1347-1359. [doi: 10.1111/jan.13971] [Medline: 30740767]

42. Myhre EL, Garnweidner-Holme L, Dahl B, Reigstad MM, Lukasse M. Development of and experiences with an informational website on early labor: Qualitative user involvement study. JMIR Form Res 2021 Oct 27;5(9):e28698 [FREE Full text] [doi: 10.2196/28698] [Medline: 34569940]
43. Rickwood D, Wallace A, Kennedy V, O'Sullivan S, Telford N, Leicester S. Young People's Satisfaction With the Online Mental Health Service eheadspace: Development and Implementation of a Service Satisfaction Measure. JMIR Ment Health 2019 May 17;6(4):e12169 [FREE Full text] [doi: 10.2196/12169] [Medline: 30994470]
44. Flavián C, Guinalíu M, Gurrea R. The role played by perceived usability, satisfaction and consumer trust on website loyalty. Information & Management 2006 Jan;43(1):1-14. [doi: 10.1016/j.im.2005.01.002]
45. Zhang MW, Ho RC, Loh A, Wing T, Wynne O, Chan SWC, et al. Current status of postnatal depression smartphone applications available on application stores: an information quality analysis. BMJ Open 2017 Dec 14;7(11):e015655 [FREE Full text] [doi: 10.1136/bmjopen-2016-015655] [Medline: 29138195]
46. Kirby PL, Reynolds KA, Walker JR, Furer P, Pryor TAM. Evaluating the quality of perinatal anxiety information available online. Arch Womens Ment Health 2018 Dec 22;21(6):813-820. [doi: 10.1007/s00737-018-0875-5] [Medline: 29931445]
47. Andrade C. The limitations of online surveys. Indian J Psychol Med 2020 Dec 13;42(6):575-576 [FREE Full text] [doi: 10.1177/0253717620957496] [Medline: 33354086]
48. Alshenqeeti H. Interviewing as a data collection method: A critical review. ELR 2014 Mar 29;3(1):39-45. [doi: 10.5430/elr.v3n1p39]

Abbreviations

BC: British Columbia
PPD: postpartum depression
WDAC: Web Development Advisory Committee

Edited by T Leung; submitted 06.09.21; peer-reviewed by A Barrera, S Shorey, L Walekr; comments to author 04.01.22; revised version received 24.04.22; accepted 30.04.22; published 23.06.22.

Please cite as:
Siddhpuria S, Breau G, Lackie ME, Lavery BM, Ryan D, Shulman B, Kennedy AL, Brotto LA
Women's Preferences and Design Recommendations for a Postpartum Depression Psychoeducation Intervention: User Involvement Study
JMIR Form Res 2022;6(6):e33411
URL: https://formative.jmir.org/2022/6/e33411
doi:10.2196/33411
PMID:35737435

Original Paper

Co-created Mobile Apps for Palliative Care Using Community-Partnered Participatory Research: Development and Usability Study

Jafar Al-Mondhiry[1], MA, MD; Sarah D'Ambruoso[2], MSN; Christopher Pietras[3], MD; Thomas Strouse[4], MD; Dikla Benzeevi[5], BS; Armen C Arevian[6], MD, PhD; Kenneth B Wells[4], MPH, MD

[1]Division of Medical Oncology, Department of Medicine, University of Southern California, Keck School of Medicine, Los Angeles, CA, United States

[2]Division of Hematology & Oncology, Department of Medicine, University of California, Los Angeles, Los Angeles, CA, United States

[3]Palliative Care Program, Department of Medicine, University of California, Los Angeles, Los Angeles, CA, United States

[4]Department of Psychiatry and Biobehavioral Sciences, UCLA Jane and Terry Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, Los Angeles, CA, United States

[5]UCLA Clinical and Translational Science Institute, Los Angeles, CA, United States

[6]Chorus Innovations, Los Angeles, CA, United States

Corresponding Author:
Jafar Al-Mondhiry, MA, MD
Division of Medical Oncology
Department of Medicine
University of Southern California, Keck School of Medicine
1441 Eastlake Ave., NTT 3440
Los Angeles, CA, 90089
United States
Phone: 1 717 919 2250
Fax: 1 717 971 4409
Email: Jafar.AlMondhiry@med.usc.edu

Abstract

Background: Open design formats for mobile apps help clinicians and stakeholders bring their needs to direct, co-creative solutions. Palliative care for patients with advanced cancers requires intensive monitoring and support and remains an area in high need for innovation.

Objective: This study aims to use community-partnered participatory research to co-design and pretest a mobile app that focuses on palliative care priorities of clinicians and patients with advanced cancer.

Methods: In-person and teleconference workshops were held with patient and family stakeholders, researchers, and clinicians in palliative care and oncology. Question prompts, written feedback, semistructured interviews, and facilitated group discussions identified the core palliative care needs. Using Chorus, a no-code app-building platform, a mobile app was co-designed with the stakeholders. A pretest with 11 patients was conducted, with semistructured interviews of clinician and patient users for feedback.

Results: Key themes identified from the focus groups included needs for patient advocacy and encouragement, access to vetted information, patient-clinician communication support, and symptom management. The initial prototype, *My Wellness App*, contained a weekly wellness journal to track patient-reported symptoms, goals, and medication use; information on self-management of symptoms; community resources; and patient and caregiver testimonial videos. Initial pretesting identified value in app-based communication for clinicians, patients, and caregivers, with suggestions for improving user interface, feedback and presentation of symptom reports, and gamification and staff coordinators to support patient app engagement.

Conclusions: The development of a mobile app using community-partnered participatory research is a low-technology and feasible intervention for palliative care. Iterative redesign and user interface expertise may improve implementation.

(JMIR Form Res 2022;6(6):e33849) doi:10.2196/33849

KEYWORDS

mobile phone; mobile apps; mobile health; mHealth; eHealth; digital health; palliative care; quality of life; survivorship; patient advocacy; oncology; patient-reported outcomes; PRO; community-partnered participatory research; CPPR

Introduction

Over the last decade, mobile app technologies for health tracking and support have become widely popular. Most available products come from consumer software companies, with only a minority being generated by health care professionals or research institutions [1]. Few studies have focused on the palliative care needs of patients with advanced cancer, despite high levels of uncontrolled symptoms, including depression, fatigue, pain, anxiety, and distress [2-4]. These needs are compounded by communication issues between patients with cancer and their treating clinicians [4,5]. Given the high cost and health system use inherent to this population [6], innovative technological solutions may help address the care needs that emerge for such patients in ambulatory settings.

Web-based interventions have shown promise in facilitating the self-management of cancer-related symptoms and communication with health care providers [7]. Specifically, the use of web-based platforms to collect patient-reported outcomes (PROs) has been associated with improvements in health-related quality of life and overall survival [8], mediated by proactive management of emergent symptoms. Efforts to integrate artificial intelligence into the design of PRO-collecting apps have resulted in improved cancer-related pain control and fewer pain-related hospital admissions [9].

To date, we are aware of no mobile apps focusing on the needs of patients diagnosed with cancer that have used community-partnered participatory research (CPPR) methods in their development. CPPR is a variant of community-based participatory research that promotes 2-way knowledge exchange and equal transfer of expertise and power sharing with the development of trust among patients, communities, health systems, policy leaders, and academic partners in the planning, process, and products of research [10,11]. CPPR highlights both community and academic shared perspectives, as distinct from community-based participatory research, which primarily focuses on academics supporting the priorities of communities; however, both focus on the importance of collaboration in authentic partnerships [11]. CPPR has been applied across diverse health and social conditions, particularly in underresourced communities, and has been used as the basis for community-level collaborative care interventions in mental health, with evidence of long-term effectiveness relative to standard individual agency training [12-14]. The extensions of this work have supported the collaborative participatory development of mental well-being support apps [15]. Such methods are rooted in the philosophy that the inclusion of patients in the development of interventions aimed at their care may increase perceptions of autonomy and competence in receiving that care, qualities associated with greater medication adherence [16], satisfaction [17], and health-related behavior changes [18,19]. For palliative care populations in particular, such inclusion may enhance a sense of dignity, a core feature of well-being threatened by advanced illness [20], and help normalize palliative care as a core component of health care that merits access to information and support [21]. This quality improvement (QI) initiative explores the experience of using CPPR methods to co-create (phase 1) and pretest (phase 2) a mobile app to meet the palliative care priorities of clinicians and patients with advanced cancer, with a focus on feasibility (inclusion of stakeholders and ease of use) and acceptability (fit with priorities of patients and providers).

Methods

Ethics Approval

The University of California, Los Angeles, Institutional Review Board provided expedited review and approval of this QI initiative (phase 1: UCLA#17-000294; phase 2: UCLA IRB#20-002047). Participation in the study was voluntary. Informed consent was obtained from participants at all levels, including focus work group sessions and the pretest study.

Setting

Study activities were conducted at an academic-community partnership in West Los Angeles, California.

Phase 1: Participatory App Development

CPPR Structure, Planning Committee, and Stakeholder Identification

CPPR uses a *Vision, Valley, Victory* (planning, implementation, and products) process guided by core principles (trust development, 2-way knowledge exchange, respect, partnered development, and equity focus) applied through a structure with a leadership council, stakeholder working groups, and broader input acquired through evaluation. This structure is similar to the *Plan-Do-Study-Act* structure for QI [22], with the distinction that CPPR is driven by iterative stakeholder feedback in every phase of development. In CPPR, the Vision stage can be its own project, including piloting, evaluation, and initial product in preparation for subsequent adaptation and main implementation [10]. Herein, we describe an initial *Vision* phase of development, including the planning, pretesting, and evaluation of an app prototype, created in preparation for a larger implementation initiative. For this project, the health system palliative care QI leadership group invited clinicians from palliative care, oncology, psychiatry, primary care, and urology, in addition to representatives from pharmacy, social work, and health information technology, to join with palliative care patients and family stakeholders to collaborate and explore options to enhance palliative care services with digital technology. Oncology faculty members from the health system outside this leadership committee were recruited to create a core group of 5 physicians and nurse practitioners comprising the provider work group. Using flyers, emails, and direct patient outreach, 4 patients receiving oncology and palliative care services in the same health system were recruited for the patient work group. A fifth patient representative was recruited as a diversity leader

with a family history of advanced cancer and palliative care exposure within the same health system.

Participatory Technology Development Platform

The goal of this planning QI (*Vision*) initiative was to collaborate with provider and patient-caregiver working groups to develop a mobile app to address the diverse needs of clinician and patient stakeholders in the delivery of high-quality palliative care for patients with advanced cancer. Chorus is a no-code app-building platform that enables individuals without technical training to use a simple, visual web interface to create interactive web-based apps optimized for mobile app use, accessible by computers or smartphones [15]. Consistent with the principles of CPPR, this allows both patients and clinicians to be involved in all aspects of product development, with previous success documented using this platform in ethnically diverse urban populations [15]. This technology allows users to create, test, and modify mobile app content in real time, with no programming experience required.

Work Group Structure and Feedback

From June to December 2017 (Figure 1), weekly provider and patient work group meetings were conducted for a total of 4 sessions per work group. Work groups were facilitated by 2 members of the QI leadership team (AA and KB). Following CPPR principles [10], the work group participants were given an orientation in the CPPR methods and Chorus app features and asked to share their experiences and perspectives. Question prompts were offered to elicit discussion. Participatory development with Chorus involved an iterative cycle of four steps: (1) identifying key barriers and opportunities for improving palliative care delivery using a mobile app tool, (2) generation of app content, (3) creating working prototypes of the mobile tool in real time by QI leads using Chorus, and (4) testing draft mobile apps in real time during workshops with iterative improvement of tools based on feedback from the work group (Figure 2). Modifications in app development continued until a consensus was reached among the stakeholders. Consistent with the principle of 2-way communication and knowledge exchange, a patient stakeholder participated in the provider work group and vice versa. Standardized reflective discussion prompts were administered at the end of each work group session to prompt discussion of process and progress and to inform agendas for subsequent sessions (Multimedia Appendix 1). Minutes were taken by the session leaders and support staff. For each work group (provider and patient-caregiver), one main overview session for each work group before the main app development was audio recorded and transcribed for subsequent qualitative analysis to illustrate the work group process. To facilitate feedback effectively and efficiently, we used rapid analysis techniques [23] to synthesize themes from work groups into generalized categories supported by representative quotes from transcripts or meeting notes. Notes, audio recordings, and transcripts from each work group were reviewed by 2 members (JA and KW) to reach an agreement on concepts and themes and select representative examples. Patient and family member participants were offered US $20 gift cards for taking part in each of the 2-hour work group sessions, in addition to parking validation.

Figure 1. Timeline for app development.

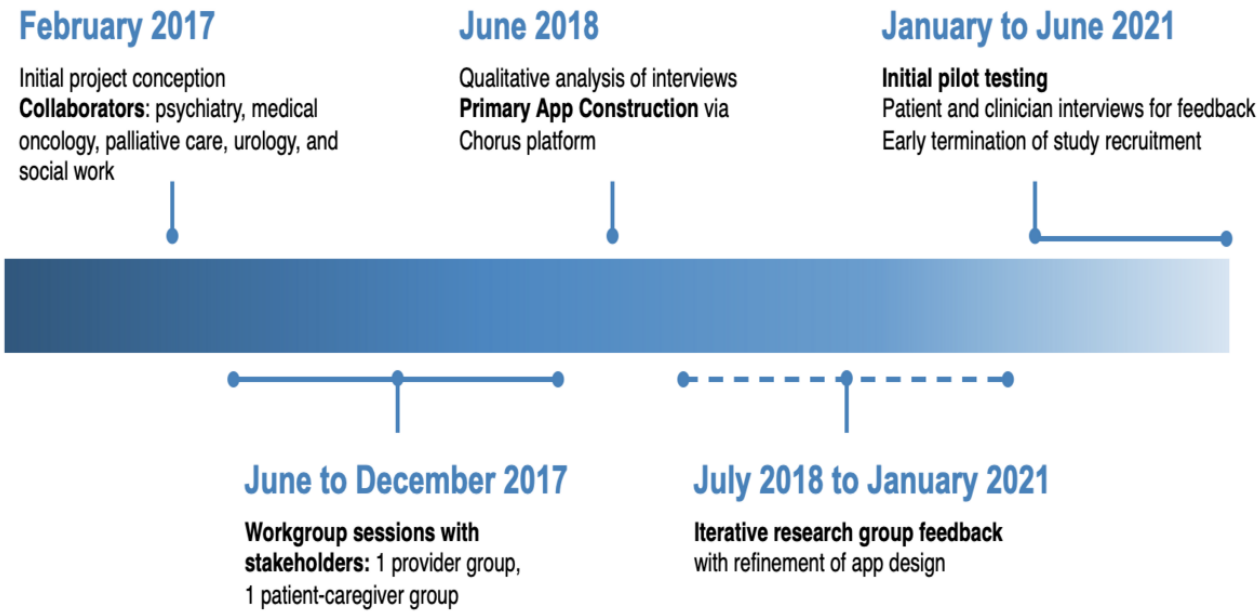
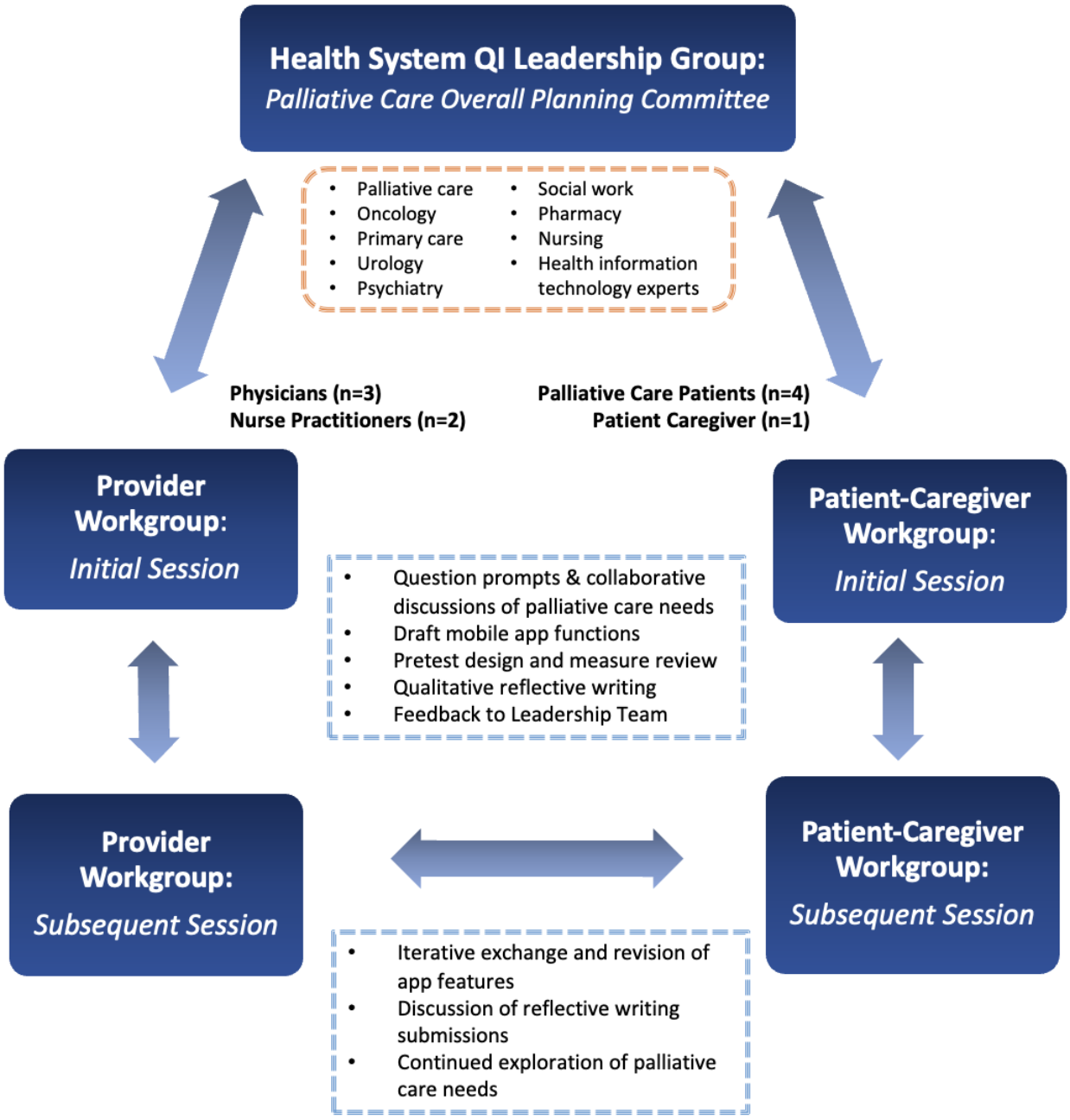


Figure 2. Work group structure and feedback cycles. QI: quality improvement.



Phase 2: Pretest

To determine the feasibility and acceptability of the initial product, the *My Wellness App*, we conducted a pretest, with the goal of recruiting at least 10 patients via a blanket email invitation to the entire palliative care patient panel of one of the authors (SD). All recruited patients were undergoing care for metastatic cancer, were aged at least 18 years, had English-language proficiency, and had access to email and a mobile smartphone. To elicit views independent of the developers, pretest participants were not involved in designing the app. Baseline surveys were administered to measure digital health preferences (developed by the work groups), a Brief Pain Inventory-Interference scale [24], and abbreviated Patient Empowerment Scales [25] (Multimedia Appendix 2). Consistent with CPPR, all scales were chosen with input from patient and provider stakeholders, with the goal of having standardized measures relevant to both parties. A provider portal was developed to allow mutual access to content shared on the app by patients with their treating clinician. The consenting clinician from the planning group provided app orientation to the participating patients, encouraged weekly *Wellness Journal* entries, and reviewed patient input via the shared provider portal during regularly scheduled clinic visits or by telephone, as indicated. Semistructured interviews were conducted and

recorded with the treating clinician (SD) and a patient to elicit feedback for improvement.

Results

Phase 1: Participatory App Development

Patient Work Groups

Drawing from their own experiences, the patients identified several goals for an app to support palliative care delivery (Table 1). One key theme mentioned during work group discussions was the difficulty in finding well-vetted material regarding their medical problems and symptoms; for example, "I think the information is out there, but it is in a sea of misinformation." Another key theme was the need for a balance between having more information on medications, supplements, herbs, drug interactions, and side effects with the need for streamlined and easy-to-read information. The third theme was the elusive nature of symptom management, with priorities shifting from day to day and difficulty knowing how to communicate with their care team appropriately. Patients sympathized with the needs of their clinicians, wanting to provide accurate and well-synthesized information; for example: "Every individual has their own reference of how serious the side effect is, and not everyone expresses it in the technical language that would be most useful for a provider." In addition to offering information to their providers, a fourth theme was that patients acknowledged that tracking symptom scores could have a personal benefit in helping to reflect on changes, better understand symptoms, and potentially know when to act. The fifth theme was placing symptoms in a functional context (ie, what they mean and how to moderate them) to help patients both demystify and get perspective on the impact of their symptoms and through that build confidence and self-efficacy in symptom management.

Table 1. Themes from patient work group sessions on app development

Theme	Example
Peer-to-peer descriptions of palliative care (dispelling misconceptions)	• "The oncologists don't necessarily communicate [about palliative care] and then people get really scared of it. I've seen it, a lot of friends of mine who are newly diagnosed and I'm trying to guide them as someone who's been sick for so long." • "My reaction was I don't want to do palliative care ... my initial reaction was this is end of life. And I didn't understand what it was at all."
Tips in understanding and managing symptoms	• "If I had put [my symptoms] into the system, and then I could see a timeline, a graph of it, I may have gone in sooner knowing, well I'm fooling myself. I've had, you know, 15 days of severe pain. I need to do something about it." • "It's so important, I mean, managing side effects enables you to get more treatment."
Tracking and encouraging progress	• "'Thriversh ip' in my community, which is the metastatic breast cancer community, it's very crucial word because what we read out there is very depressing and a downer."
Improved patient-clinician communication	• "If we're going to use pain scales, put it in context. Because I know sometimes when I'm asked, 'What's your pain?,' I feel a lot of pain. But if I really think about it, you know, I drove today, so I'm in pain, but I'm able to drive or I walk the dog. I'm like, yeah, I didn't feel good but I was able to walk the dog, so it's not that horrific." • "Every individual has their own reference of how serious the side effect is, and not everyone expresses it in the technical language that would be most useful for a provider."
Building confidence	• "It's a bit of perspective. So having those kinds of trigger questions [around symptom context and daily activities] may make patients to refocus on what's good."
Vetted information about medications and herbs	• "I think the information is out there, but it is in a sea of misinformation." • "People are out there and they're Googling and looking for that, and they're getting scared. And many of these [health] systems don't have an official list of our recommended resources."
End-of-life care planning	• "I think that a way to help patients get through that [end-of-life planning] process may be information. It's very complicated." • "I think that's such an important issue that people are so scared of, because it causes all sorts of family fighting ... maybe just having an advance directive tool up on the site, having that paperwork available."
Improved patient advocacy	• "Explaining why [you should] have someone with you, why record your [clinic] sessions, why bring a list of questions, you could say that studies show that the recall rate after leaving a doctor's office is at best 30%."

In addition, although patients in this work group were engaged in palliative care, they were all aware of the stigma surrounding palliative care as a term often misunderstood to describe only end-of-life care or hospice care, rather than surviving and thriving despite serious illness. As a result of this input, suggestions were made for the app to address a more expansive definition of palliative care as a positive service that provides supportive care and symptom or side effect management from diagnosis onward, highlighting the benefits of palliative care in improving patient quality of life while in active treatment. To proactively address stigma and normalize the delivery of palliative care, QI leaders and patients collaboratively suggested including peer-to-peer engagement in the form of video testimonials on the app. Empathizing with the needs of other

patients recently diagnosed with cancer, patient stakeholders wanted an app that addressed challenges in survivorship, such as how to have successful encounters with their clinicians, with advice delivered to app users in laymen's terms. Videos from patient peers explaining how to advocate for oneself, how to recruit help from caregivers during appointments, and what to expect from palliative care were suggested as solutions. In addition, patients requested vetted lists of resources to help them quickly navigate their changing needs and help with end-of-life and advance care planning. Ultimately, they felt the app features needed to be rich and adaptable to each patient's needs, with opportunities for growth and customization: "Something that I learned being sick over these last three years is that every single person, even with the same diagnosis handles it differently." Stakeholders requested that the videos would represent diverse populations and experiences with palliative care, including clinicians, patients, and family members.

Provider Work Groups

Several consistent themes and goals were developed over the course of the provider work group sessions (Table 2). Concerns around app time demands from patients and providers were balanced with the hope of creating greater efficiency in patient-provider communication. Providers acknowledged the difficulty patients had in discussing symptoms and worked to create app functions that measure symptom severity, track progress, and alert providers when threshold symptom levels are met. The goal of improving communication was the dominant theme, using the app to streamline patient symptom trends between visits and triage emergencies. A specific focus was placed on pain as a key symptom affecting the quality of life and the creation of a comprehensive pain diary to allow patient reflection on aggravating or alleviating factors and context. Providers were interested in how the app could elicit a patient's medication use patterns, promote adherence, and provide a way to review and address symptoms. Providers were sympathetic to the emotional and socioeconomic impact of cancer and hoped to provide patients with tools to address these barriers to care through encouragement and a curated list of local resources. The app was suggested as a potential anchoring point for caregivers, allowing them access to symptom reporting on behalf of the patient and symptom management tips, features that may ease their anxiety while advancing the patient's care. This was also seen as a support for patients with advanced symptoms who may not have the capacity to manage their own needs or communicate effectively with their clinician.

Table 2. Themes from provider work group sessions on app development

Theme	Example
Comprehensive pain diary	• "If you're going to do the pain score, tracking, well, when the pain was this—what did I do for it? Or what did I take?"
Medication reminders and use monitoring	• "If you're working with a patient, you can say, well, how often do you feel you need a reminder? And then do you have the flexibility of every day, every other day?"
Provider alert messaging	• "If it [patient pain level] meets a certain threshold, then you know, then there's a prompt to send out an alert."
Improved triage and communication	• "The two things I hear from people all the time is: it's hard for them to describe things, and they don't know what to do when to call, when to panic. And if you address those two—you say, okay, if your pain gets to a five, I need you to pick up the phone—then they relax." • "Because then when they call, you're like, oh, let me have a look [at patient generated symptom reports on the app] and see what you've done. Ok, that didn't work. Ok, that worked. Oh, I can try this." • "If we're going to change somebody's pain medication, it would be great to know what the last four or five days looked like."
Addressing barriers to care	• "Resources that are kind of all scattered across the internet, consolidating those resources in one place and have those be resources that are vetted." • "We can broaden the application so that not only do we manage symptoms that are disease related, but also any impediments to the delivery of care. That's really important because I mean, to get people to comply with what you propose..." • "We probably should include some kind of either logistical or financial scale, because I think sometimes those issues contribute to compliance and symptom management, but patients are not prompted and won't volunteer that."
Encouraging and supporting patients through difficulty	• "I think when I first get someone who's newly metastatic, it's a whole shift in psyche ... And when they progress on their therapy, even though they've been through it once they're like, oh crap, I'm back at the beginning again. And so how do we get them moving through that next set of events and getting that [next] therapy started. Because they're sort of on this constant roller coaster, emotional and physical and psychological. Living scan to scan."
Vetted educational material for patients	• "We're always telling patients, 'don't look at that—look at this.'"
Partnering with caregivers	• "By the time they're at that point [of serious symptoms], a lot of times the caregiver is involved. And I think the caregivers really calm down if you give them stuff to work from, like a document. I think that's really important because most of them feel so inadequate."

App Design

On the basis of iterative stakeholder input and review, the initial *My Wellness App* prototype contained four core features: (1) a *Wellness Journal*, where patients are encouraged to make weekly entries; (2) *Tips & Tools* for symptom and medication self-management; (3) *Voices of Palliative Care*, where brief videos of palliative care clinicians, patients, and caregivers can be viewed; and (4) a *Resources List*, where a directory of local patient support resources can be viewed, including links and information to access transportation, home services, hospice, and legal and insurance benefits. To enhance patient and clinician communication, the *Wellness Journal* contained a modified Edmonton Symptom Assessment Scale (ESAS) [26], a body map for visualization of pain, areas to report medication use frequency, goal setting, and open diary free text entries to permit direct patient *voice* (Figure 3). ESAS entries were displayed longitudinally to track trends and were available for both patients and their treating clinicians (Figure 4). The initial app design encouraged weekly entries with automated email reminders sent to nonresponders each week. Owing to technical limitations, no integration with the existing electronic medical records or alert messaging for symptom score thresholds was included in the initial pretest phase. The *Tips & Tools* feature included brief information reviews for common issues with links to videos and more in-depth reviews and resources. Specific topics included symptom management (eg, nausea, diarrhea, and constipation), coping with difficult emotions, mindfulness, nutrition, and the use of herbs and botanicals, all areas suggested by stakeholders. The *Voices* section included members of the provider and patient work group, offering videos of patient introductions for *What is palliative care?* and other features of the app, the importance of symptom monitoring, stories of survivorship, and tips for maximizing support and clinician communication. One area of shared concern for patients and providers was addressing the stigma of palliative care through effective communication. Feedback from a patient is as follows: “Do you want to keep using the term palliative care, cause it scares people. I mean, like I said, it scared me.” An example from a patient in the provider group is as follows: “The advanced stage breast cancer survivor has a negative connotation. If you want to expand this to a greater cancer or illness community, then maybe there are other words that connect with a greater population.” A provider’s response was as follows: “So I think we have to rebrand this, so people see it as a vital tool that ensures their success during their treatment.” This shared concern led to the proposed app name, *My Wellness App*.

Figure 3. Example of My Wellness Journal details.

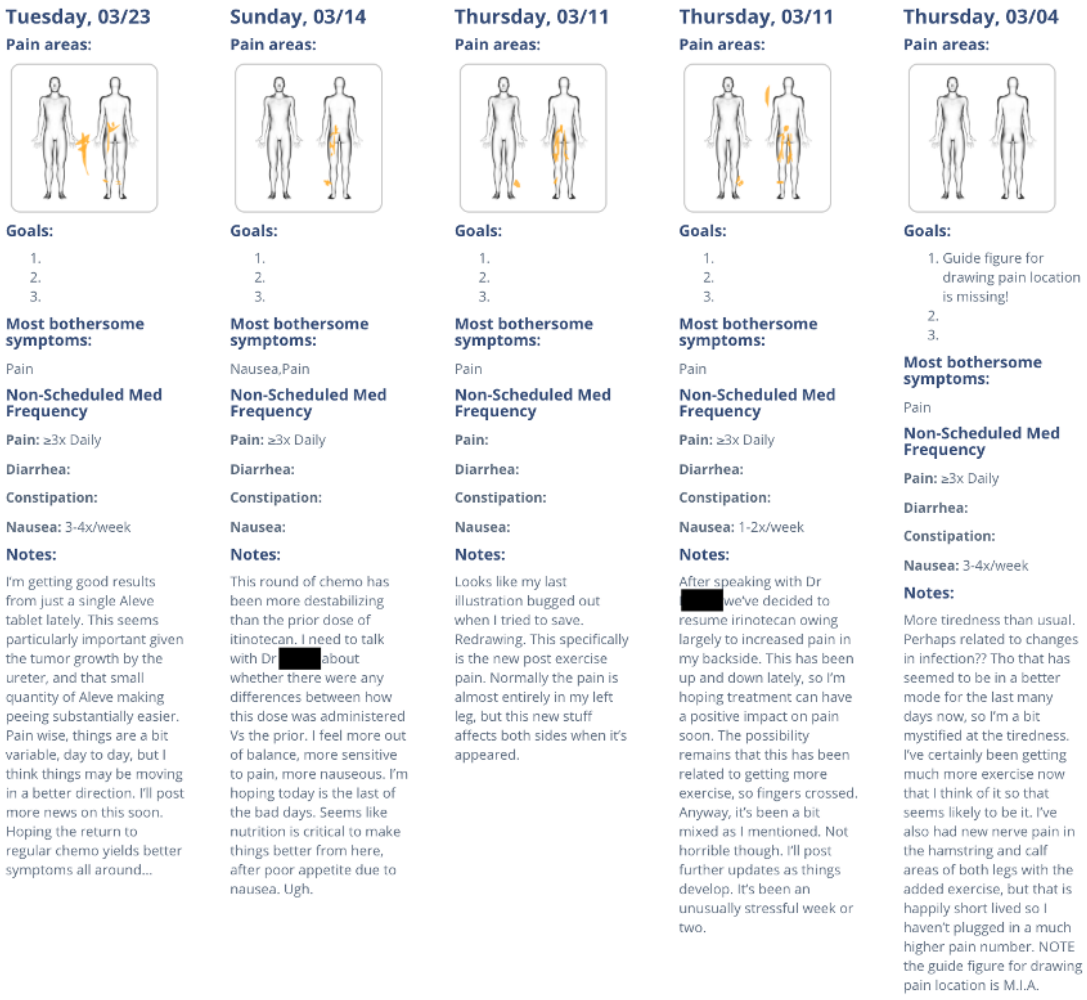
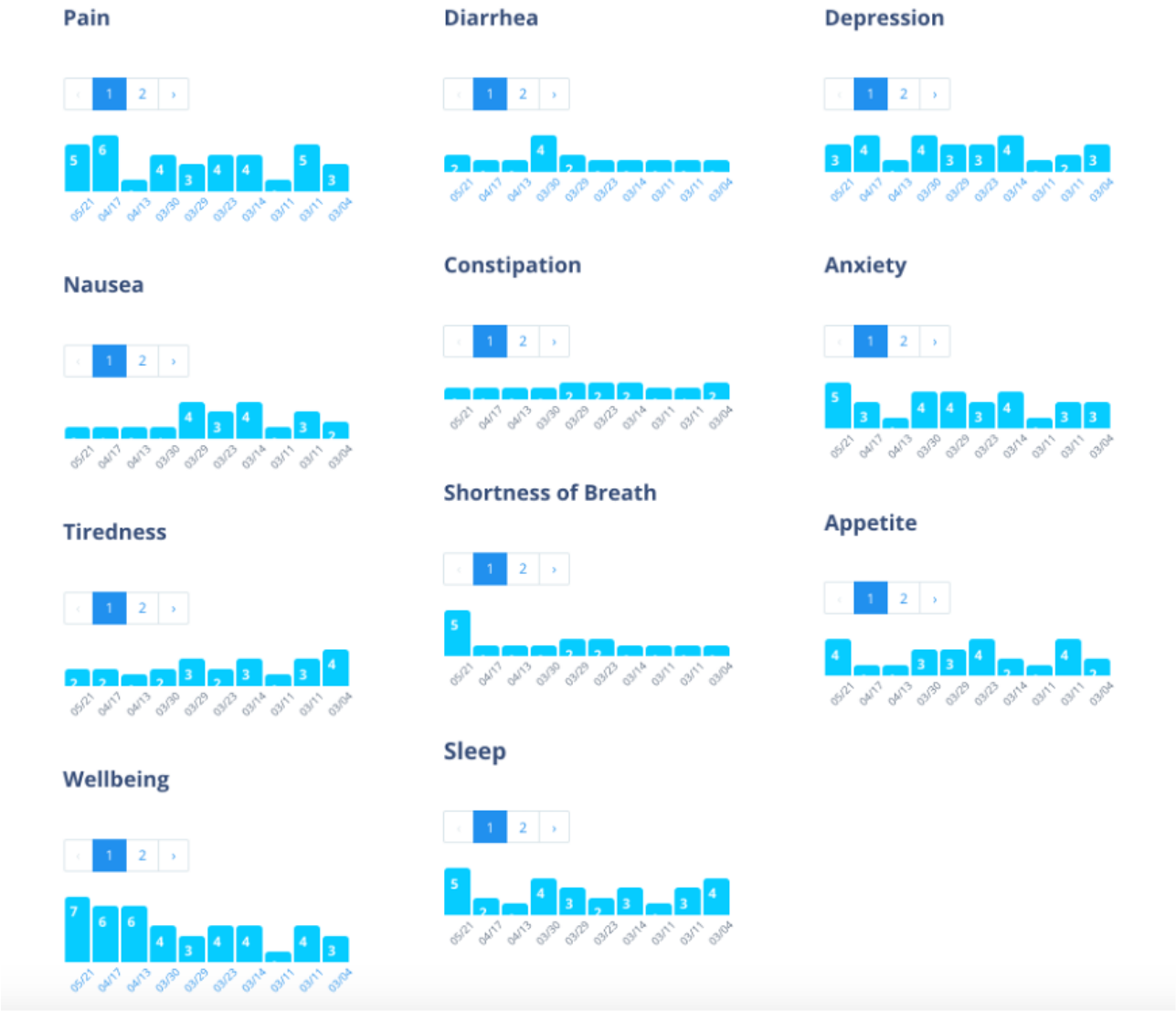


Figure 4. Example of modified Edmonton Symptom Assessment Scale representations.



CPPR Process for App Development and Pretest Design

Throughout the stakeholder work group sessions, multiple demonstrations of CPPR principles occurred [10], summarized by key themes and examples of quotes from transcripts of key work group sessions at the transition from initial planning to app drafting (Tables 3 and 4). Consistent with the principle of coequal leadership and power through 2-way knowledge exchange, the design of the final app product was equally informed by patient and provider stakeholder input. The protection and elevation of vulnerable participant voices is a core feature of CPPR; to avoid the threat of power imbalances in participation, separate work groups for providers and patients provided secure platforms for interaction in the design process. The 2-way knowledge exchange was illustrated by the rich interaction between academic leaders and individual work groups (see examples in Table 3). Additional knowledge exchange was made available by the inclusion of patients within the provider work group and vice versa. The principle of trust development was evident in the openness of patients to share vulnerable experiences about symptoms, concerns with the dying process, and conflicts experienced with health systems and providers. Providers also demonstrated trust in disclosing their frustrations with patient care and health systems. Respect was evidenced by repeated invitations for stakeholder perspectives as well as agreement and expansion by patients of academic leader suggestions (eg, creating videos). Partnered development took place through the interaction of academic leaders and patients in defining goals (such as having informational resources and access to a shared portal) and reviewing images and options for videos (recording of patient stakeholders). The principle of equity was pursued through the inclusion of a patient caregiver representing ethnic diversity, and attention was paid to assessing financial barriers. In addition, stakeholders commented on the potential needs of vulnerable populations, for example, providing navigator or caregiver support tools for older adults with technology access limitations.

Table 3. Community-partnered participatory research process in patient-caregiver work group at transition to app development.

Principle	Example (quote)
Coequal leadership and power	
• Attention to patient-provider information and power imbalances (eg, creation of platforms for app design and engagement) • Input of patient and family stakeholders on app design features	• "I think it would be helpful to have a way to get information about your diagnosis and the chemo." • "[Establishing] a way to email a doctor where you don't have to know his email address."
Two-way knowledge exchange	
• Rich interaction of academic leaders and patients • Inclusion of patients in provider work groups (and vice versa)	• Patient: "The advanced stage breast cancer survivor has a negative connotation." • Provider Response: "So I think we have to rebrand this, so people see it as a vital tool that ensures their success during their treatment."
Trust development	
• Frequent exchange of vulnerable experiences • Disclosure of conflicts with health system and providers by patient-caregiver stakeholders	• "It's not that you're dying. You're about to die. So you need to have an advanced directive. People need to know what your wishes are." • "I had a really painful morning. And then I did my activities during the day and I forgot about the morning, That's something that the doctor needs to know. But they're not clear on that."
Respect	
• Repeated invitations for stakeholder perspectives • Agreement and expansion by stakeholders of academic leader suggestions	• Leader: "[We're] hoping we could think through how we might be able to help the experience of individuals in palliative care." • Leader suggestion: track "how's your pain going?" • Patient response: "That would be actually very helpful because you don't know what you don't. I think that's actually a really good idea."
Partnered development	
• Interaction of academic leaders and patients in defining goals (eg, creating informational resources and access to a shared portal)	• Technology lead: "Do you think you would want a summary screen?" • Patient response: "I think for me, a little journal, I keep it on my phone. If this was on the system, then I wouldn't have to think, how many days was there pain, you know, moderate pain or severe pain."
Equity	
• Inclusion of multiple patients and caregiver with racial and ethnic diversity • Attention to potential disparities	• "I learned being sick over these last years that every single person, even with the same diagnosis handles it differently."

Table 4. Community-partnered participatory research process in provider work group at transition to app development.

Principle	Example (quote)
Coequal leadership and power	
• Input of provider stakeholders on app design features • Attention to patient-provider power imbalances (eg, creation of platforms for app design and engagement)	• Provider: "If a patient has pain and they're on pain medication, they could potentially be inputting their pain symptoms on a daily basis into this web app. And that gets fed back to the clinician at a regular visit."
Two-way knowledge exchange	
• Rich interaction of academic leaders and providers • Inclusion of patients in provider work groups (and vice versa)	• Provider: "The other thing I would love is a, a calendar that records things for the patients, because we constantly use that as a way to track therapies as well." • Technology response: "Manual would be a way to do that, to see if it's helpful and if it works and we can sort out to automate."
Trust development	
• Frequent exchange of vulnerable experiences • Disclosure of conflicts with health system and patients by each provider stakeholders	• "I don't know if I can do that for every patient. Do we expect them to actually be logging into the website?" • "Physician or practitioner who is having to go into their inbox multiple times a day, dealing with everything that comes through and there's no priority currently."
Respect	
• Repeated invitations for stakeholder perspectives • Agreement and expansion by stakeholders of academic leader suggestions	• Technology lead: "Is it possible to help improve aspects around palliative care with technologies like apps, could they be tailored and created in a way that might address problems that we are having either as providers or from the patient's perspective or caregivers of palliative care?" • Provider: "I mean, I welcome more data rather than less. As long as they can do it."
Partnered development	
• Interaction of academic leaders and providers and patients in defining goals (eg, creating informational resources and access to a shared portal)	• Provider: "I think when I first get someone who's newly metastatic, cause it's a whole shift in psyche. And how you treat them, getting started on new therapy. And even when they progress, even though they've been through it, I'm back at the beginning again. And so how do we get them moving through that next set of events and getting therapy." • Technology response: "If you're working with a patient, you can say, well, how often do you feel you need a reminder? And then do you have the flexibility of every day, every other day? And if they are going to forget, you need a family reminder."
Equity	
• Inclusion of patient caregivers and members representing ethnic diversity • Attention to potential disparities	• "We probably should include some kind of either logistical or financial scale, because sometimes those issues contribute to compliance and simply naturally, but patients are not prompted well, to be able to afford their treatment." • "One of the things we have to be careful because if we don't we create the system that others cannot access and we are inadvertently discriminatory."

Similarly, CPPR principles were followed in developing the pretest study by reviewing plans and gathering input studies from the patient-caregiver and provider work groups. An example of enthusiasm from a patient stakeholder is as follows: "I think you'll find a lot in the beta testing as you get information back." Providers considered who should be included in the pilot, after a QI lead asked the following question: "Is it people who are relatively early with minimal symptoms, or later on, more advanced?" The provider stakeholder response was as follows: "I think you could get both. I mean, I have some patients that come in newly metastatic or in acute symptom crisis until we can get their disease under control that, and then we have some that are asymptomatic with their disease." Stakeholders also gave input on the selection of items and measures.

Phase 2: Pretest Findings

A total of 11 patients completed the consent forms and were registered with the app, representing a diversity of cancer diagnoses (n=4, 36% breast cancer; n=3, 27% lung; n=2, 18% colorectal; and n=2, 18% other) and ages (median 58, range 49-82 years; Multimedia Appendix 3). Of these patients, 9 (82%) completed the baseline surveys. The patients identified predominantly as "Caucasian" (8/9, 89%), female (6/9, 63%), held either a 4-year (5/9, 56%) or postgraduate degree (3/9, 33%), and had high composite pain scores on the Brief Pain Inventory-Interference scale (average 6.54). Over the first 6

months of the pretest study activity, 3 patients regularly participated in weekly journal entries, with the other 6 completing one or none during the entire study period. The average composite pain scores between high app users and nonusers were comparable (6.42 vs 6.59). Feedback in the form of semistructured interviews with the treating clinician for all study patients, as well as with one regular user of the app, revealed several key insights (Textbox 1). Overall, it was noted that the completion of surveys, journal entries, and interviews was limited by the high symptom burden of patients, with 2 participating patients dying shortly after the study period.

Textbox 1. Themes from clinician and patient feedback during pretest study.

Clinician feedback

- Value in body mapping
- Lack of app navigator support
- Response variability (differences in numeric scale reporting)
- Opportunities for caregivers
- Higher symptom burden: higher benefit from app use

Patient feedback

- Sense of preparedness in symptom communication
- Limited patient onboarding (confusion about app functions)
- Lack of feedback and reinforcement
- Desire for improved user interface
- Log-in friction (absence of home screen app icon)

From the patient's perspective, a lack of formal onboarding to the app left them unaware of many core features, including goal setting and tips for self-management. Patients expressed a desire for a better app interface, specifically for data representation and interpretation of those data in terms of important trends or changes, as well as symptom management suggestions. A user derived motivation to use the app because of the connection it brought them with their clinician and that participation via the app made them a *good citizen* of the clinic: "I like the promise of it ... and knowing the right information to talk about when we meet." For most, however, motivation was a challenge, citing a lack of feedback and encouragement from the app itself to continue inputting data. Suggested solutions included gamification of inputs, with different milestones for completion, coupled with haptics upon each submission. As a web-based app formatted for mobile phone use, the app itself was not one that users could find in an app store or have loaded as an icon on their home screen, a detail that bothered some users and created an additional point of friction in accessing the app more quickly or regularly. Using this more familiar format was suggested as having potential for creating push notifications and visual cues over the icon to remind users of outstanding tasks to complete as part of being a patient user.

From the treating clinician's perspective, the lack of a patient navigator created several challenges. First, it increased the amount of time required by the provider to orient each patient to the app and its features. Second, it placed the onus on the clinician to expeditiously review and respond to all *Wellness Journal* entries that might signal uncontrolled symptoms. Creating threshold symptom level scores for automated messaging to the clinician was proposed as a potential solution; however, at a minimum, this would create an added layer of work for the clinician, and significant variability in symptom reporting for each patient made this potentially challenging (eg, some patients regularly report pain at an 8 out of 10, requiring frequent responses). Similarly, providing personalized reminders or check-ins that could increase motivation and adherence was left to the treating clinician.

App features that were of particular value included the body map, which helped quickly identify changes in the nature of pain symptoms and provided a meaningful jump-off for clinical exploration. Trends in app use and value to the clinician seemed to correlate with patients who had a higher symptom burden from disease activity, changes in treatment, or both. Finally, as predicted in the work groups, caregivers of patients enrolled in the study found value in filling out *Wellness Journal* entries with the patient, both as a way to formally check in with one another and to become more actively connected to the patient's care by communicating with their clinician.

Discussion

Principal Findings

Herein, we describe the first known report of applying CPPR structure and principles to co-create (phase 1) and pretest (phase 2) a mobile app based on patient and clinician stakeholder needs for palliative care delivery, as part of the *Plan* phase of QI initiative and *Vision* phase of CPPR [10,22]. Several examples of web-based and mobile technology-based interventions to elicit PROs and improve patient-provider communication exist in the literature [27-35]. However, in this study, we describe a process of identifying and addressing needs at a local level through a QI initiative, with stakeholders co-leading the app design and pretest process, following CPPR principles of trust, respect, 2-way knowledge exchange, and coleadership [10].

Implementation in our study also follows from the framework of the self-determination theory, which states that behavior is driven by 3 primary psychological needs: autonomy, competence, and relatedness [36]. By structuring interventions around the needs and input of local stakeholders (relatedness), promoting patient engagement in the process of weekly symptom journaling and goal setting (competence), and eliciting feedback for immediate app improvements and tailoring to local group needs (autonomy), we believe our study followed this model in conjunction with CPPR. However, the ultimate goal of patient engagement is to activate health-related changes and participation in the process of medical care, goals reserved in this case for a follow-up implementation and evaluation process (*Valley*) informed by suggestions for app design improvements from the initial planning (*Vision*) phase. Pretesting suggested the feasibility of some engagement in app use while revealing the need to enhance design and engagement, particularly given the high clinical needs in the patient population as well as the time limits of providers.

Accordingly, the next steps for development may include the incorporation of a patient navigator to assist in the management of patient-generated data and improve the process of providing timely and personalized follow-up for changing symptomatology, while reducing provider time burden in explaining and monitoring app engagement. Data summaries and suggestions for clinicians and patients in the management of symptoms could be enhanced using artificial intelligence to generate clinician support recommendations, as demonstrated in previous studies [9,30]. This may serve the expressed desire from patients for increased communication from the app, providing cues for continued participation and streamline information review for clinicians. Expansion of recruitment to include multiple clinicians and a larger cohort of patients at the next stage could provide additional iterative feedback for app redesigns as part of the 2-way knowledge exchange [10].

We believe our QI planning (*Vision*) initiative and pretest suggest that the co-creation of a mobile app care is a feasible, low-technology strategy for potentially improving the delivery of palliative care while lowering the bar for patient and clinician participation in such initiatives. However, our pretest suggests that such app products require continued improvements, creativity, and skills in app interface design to enhance patient motivation and support participation. Although not specifically measured in this study, we may borrow the language of the Technology Acceptance Model framework [37] in describing that clinician and patient app users in phase 2 identified significant perceived usefulness for this app but also pointed out areas where perceived ease of use could be improved. Perhaps out of necessity, most patients with advanced cancer receiving palliative care have already established other mechanisms for symptom reporting and clinic contact (eg, telephone, email, and patient portal messaging). New technologies, such as apps, need to offer even greater convenience and user gratification than these currently available channels, which may then engage patients to explore other content included from stakeholder input. Gamification, haptics, and smartphone icon interfaces are examples of features that may not be immediately recognized or implemented by stakeholders unfamiliar with app design; however, such features may significantly complement and motivate the app experience. Implementation at this level will require additional partnership with technology leaders familiar with these forms of app development, coupled with orientation and training of patient and provider stakeholders to co-create these features, an approach similar to that used in developing a COVID-19 wellness website integrating stakeholders and technology input [38].

Limitations

The limitations of this study include the preliminary pretest phase of development, with a small number of patients enrolled from a single institution and a single coordinating provider-investigator. As a QI effort using stakeholder-partnered development and evaluation, the results are, by design, meant to be reflective of the local community and thus may not reflect the needs of other communities, regions, or health systems. Furthermore, we acknowledge that the qualitative nature of the data generated from planning and small stakeholder groups may limit generalizability, even within this community. The authors hope that future expansion into other settings and populations, as well as attending to this initial feedback, will enrich the understanding of palliative care needs for patients with advanced cancer through iterative and diverse stakeholder inputs. Additional limitations in the pretest include our focus on English-speaking patients and those with smartphone access, which potentially excludes some underresourced populations.

Conclusions

Overall, this planning initiative and pretest reinforce the feasibility of applying the CPPR framework to stakeholder co-created palliative care apps, with recommendations identified to more consistently and effectively support patient-caregiver use of the app with their clinicians.

Acknowledgments

The authors would like to acknowledge all patient and clinician stakeholders who generously offered their time to the planning phases of this study. Additional valued input was provided by the following clinicians (listed in alphabetical order): Jonathan Bergman, MD; Amy L Cummings, MD; Jennifer O'Hora; Matthew J Loscalzo, Licensed Clinical Social Worker; Anne M Walling, MD, PhD; and Neil Wenger, MD. Chorus software and project support were provided by Catherine Wilkerson, a staff research associate with the University of California, Los Angeles, Semel Institute for Neuroscience and Human Behavior.

This work was supported by funding from the Resnick Hospital Quality Team account (401299-1B-62190), Samuel Steinberg Family Foundation Palliative Care Research Endowment, Maddie Katz Endowed Chair in Palliative Care Research and Education, and David Weil Endowed Chair.

Conflicts of Interest

AA reports owning stock, patents, and leadership within Chorus Innovations, Inc, where he serves as the chief executive officer.

Multimedia Appendix 1
Work group reflection sheet.
[DOCX File , 19 KB - formative_v6i6e33849_app1.docx]

Multimedia Appendix 2
Baseline survey for pretest study participants.
[DOCX File , 31 KB - formative_v6i6e33849_app2.docx]

Multimedia Appendix 3
Pretest study patient characteristics.
[DOCX File , 40 KB - formative_v6i6e33849_app3.docx]

References

1. Tabi K, Randhawa AS, Choi F, Mithani Z, Albers F, Schnieder M, et al. Mobile apps for medication management: review and analysis. JMIR Mhealth Uhealth 2019 Sep 11;7(9):e13608 [FREE Full text] [doi: 10.2196/13608] [Medline: 31512580]
2. Miovic M, Block S. Psychiatric disorders in advanced cancer. Cancer 2007 Oct 15;110(8):1665-1676 [FREE Full text] [doi: 10.1002/cncr.22980] [Medline: 17847017]
3. Cleeland CS. The impact of pain on the patient with cancer. Cancer 1984 Dec 01;54(11 Suppl):2635-2641. [doi: 10.1002/1097-0142(19841201)54:2+<2635::aid-cncr2820541407>3.0.co;2-p] [Medline: 6498754]
4. Smith TG, Troeschel AN, Castro KM, Arora NK, Stein K, Lipscomb J, et al. Perceptions of patients with breast and colon cancer of the management of cancer-related pain, fatigue, and emotional distress in community oncology. J Clin Oncol 2019 Jul 01;37(19):1666-1676 [FREE Full text] [doi: 10.1200/JCO.18.01579] [Medline: 31100037]
5. Hack TF, Degner LF, Parker PA, SCRN Communication Team. The communication goals and needs of cancer patients: a review. Psychooncology 2005 Oct;14(10):831-847. [doi: 10.1002/pon.949] [Medline: 16200519]
6. May P, Garrido MM, Aldridge MD, Cassel JB, Kelley AS, Meier DE, et al. Prospective cohort study of hospitalized adults with advanced cancer: associations between complications, comorbidity, and utilization. J Hosp Med 2017 Jun;12(6):407-413. [doi: 10.12788/jhm.2745] [Medline: 28574529]
7. Kim AR, Park HA. Web-based self-management support interventions for cancer survivors: a systematic review and meta-analyses. Stud Health Technol Inform 2015;216:142-147. [Medline: 26262027]
8. Basch E, Deal AM, Kris MG, Scher HI, Hudis CA, Sabbatini P, et al. Symptom monitoring with patient-reported outcomes during routine cancer treatment: a randomized controlled trial. J Clin Oncol 2016 Feb 20;34(6):557-565 [FREE Full text] [doi: 10.1200/JCO.2015.63.0830] [Medline: 26644527]
9. Kamdar M, Centi AJ, Agboola S, Fischer N, Rinaldi S, Strand JJ, et al. A randomized controlled trial of a novel artificial intelligence-based smartphone application to optimize the management of cancer-related pain. J Clin Oncol 2019 May 20;37(15_suppl):11514. [doi: 10.1200/jco.2019.37.15_suppl.11514]
10. Jones L, Wells K. Strategies for academic and clinician engagement in community-participatory partnered research. JAMA 2007 Jan 24;297(4):407-410. [doi: 10.1001/jama.297.4.407] [Medline: 17244838]
11. Israel BA, Schulz AJ, Parker EA, Becker AB. Review of community-based research: assessing partnership approaches to improve public health. Annu Rev Public Health 1998;19:173-202. [doi: 10.1146/annurev.publhealth.19.1.173] [Medline: 9611617]
12. Jones L. Commentary: 25 years of community partnered participatory research. Ethn Dis 2018 Sep 6;28(Suppl 2):291-294 [FREE Full text] [doi: 10.18865/ed.28.S2.291] [Medline: 30202180]
13. Wells KB, Jones L, Chung B, Dixon EL, Tang L, Gilmore J, et al. Community-partnered cluster-randomized comparative effectiveness trial of community engagement and planning or resources for services to address depression disparities. J Gen Intern Med 2013 Oct;28(10):1268-1278 [FREE Full text] [doi: 10.1007/s11606-013-2484-3] [Medline: 23649787]
14. Arevian AC, Jones F, Tang L, Sherbourne CD, Jones L, Miranda J, Community Partners in Care Writing Group. Depression remission from community coalitions versus individual program support for services: findings from community partners in care, Los Angeles, California, 2010-2016. Am J Public Health 2019 Jun;109(S3):S205-S213. [doi: 10.2105/AJPH.2019.305082] [Medline: 31242001]
15. Arevian AC, O'Hora J, Jones F, Mango J, Jones L, Williams PG, et al. Participatory technology development to enhance community resilience. Ethn Dis 2018 Sep 6;28(Suppl 2):493-502 [FREE Full text] [doi: 10.18865/ed.28.S2.493] [Medline: 30202203]
16. Kennedy S, Goggin K, Nollen N. Adherence to HIV medications: utility of the theory of self-determination. Cogn Ther Res 2004 Oct;28(5):611-628. [doi: 10.1023/b:cotr.0000045568.95219.e2]

17. Williams GC, McGregor HA, King D, Nelson CC, Glasgow RE. Variation in perceived competence, glycemic control, and patient satisfaction: relationship to autonomy support from physicians. Patient Educ Couns 2005 Apr;57(1):39-45. [doi: 10.1016/j.pec.2004.04.001] [Medline: 15797151]
18. Williams GC, Grow VM, Freedman ZR, Ryan RM, Deci EL. Motivational predictors of weight loss and weight-loss maintenance. J Pers Soc Psychol 1996 Jan;70(1):115-126. [doi: 10.1037//0022-3514.70.1.115] [Medline: 8558405]
19. Williams GC, McGregor HA, Zeldman A, Freedman ZR, Deci EL. Testing a self-determination theory process model for promoting glycemic control through diabetes self-management. Health Psychol 2004 Jan;23(1):58-66. [doi: 10.1037/0278-6133.23.1.58] [Medline: 14756604]
20. Chochinov HM. Dignity-conserving care--a new model for palliative care: helping the patient feel valued. JAMA 2002 May 01;287(17):2253-2260. [doi: 10.1001/jama.287.17.2253] [Medline: 11980525]
21. Noble C, Grealish L, Teodorczuk A, Shanahan B, Hiremagular B, Morris J, et al. How can end of life care excellence be normalized in hospitals? Lessons from a qualitative framework study. BMC Palliat Care 2018 Aug 08;17(1):100 [FREE Full text] [doi: 10.1186/s12904-018-0353-x] [Medline: 30089484]
22. Langley G, Moen RD, Nolan KM, Norman CL, Provost LP. The Improvement Guide: A Practical Approach to Enhancing Organizational Performance. 2nd edition. San Francisco, CA, USA: Jossey-Bass Publishers; Apr 2009.
23. Gale RC, Wu J, Erhardt T, Bounthavong M, Reardon CM, Damschroder LJ, et al. Comparison of rapid vs in-depth qualitative analytic methods from a process evaluation of academic detailing in the Veterans Health Administration. Implement Sci 2019 Feb 01;14(1):11 [FREE Full text] [doi: 10.1186/s13012-019-0853-y] [Medline: 30709368]
24. Miettinen T, Kautiainen H, Mäntyselkä P, Linton SJ, Kalso E. Pain interference type and level guide the assessment process in chronic pain: categorizing pain patients entering tertiary pain treatment with the Brief Pain Inventory. PLoS One 2019 Aug 20;14(8):e0221437 [FREE Full text] [doi: 10.1371/journal.pone.0221437] [Medline: 31430355]
25. Bulsara C, Styles I, Ward AM, Bulsara M. The psychometrics of developing the patient empowerment scale. J Psychosoc Oncol 2006;24(2):1-16. [doi: 10.1300/J077v24n02_01] [Medline: 17046803]
26. Chang VT, Hwang SS, Feuerman M. Validation of the Edmonton Symptom Assessment Scale. Cancer 2000 May 01;88(9):2164-2171. [doi: 10.1002/(sici)1097-0142(20000501)88:9<2164::aid-cncr24>3.0.co;2-5] [Medline: 10813730]
27. Abernethy AP, Zafar SY, Uronis H, Wheeler JL, Coan A, Rowe K, et al. Validation of the Patient Care Monitor (Version 2.0): a review of system assessment instrument for cancer patients. J Pain Symptom Manage 2010 Oct;40(4):545-558 [FREE Full text] [doi: 10.1016/j.jpainsymman.2010.01.017] [Medline: 20579839]
28. Basch E, Artz D, Iasonos A, Speakman J, Shannon K, Lin K, et al. Evaluation of an online platform for cancer patient self-reporting of chemotherapy toxicities. J Am Med Inform Assoc 2007;14(3):264-268 [FREE Full text] [doi: 10.1197/jamia.M2177] [Medline: 17329732]
29. Berry DL, Blumenstein BA, Halpenny B, Wolpin S, Fann JR, Austin-Seymour M, et al. Enhancing patient-provider communication with the electronic self-report assessment for cancer: a randomized trial. J Clin Oncol 2011 Mar 10;29(8):1029-1035 [FREE Full text] [doi: 10.1200/JCO.2010.30.3909] [Medline: 21282548]
30. Cooley ME, Blonquist TM, Catalano PJ, Lobach DF, Halpenny B, McCorkle R, et al. Feasibility of using algorithm-based clinical decision support for symptom assessment and management in lung cancer. J Pain Symptom Manage 2015 Jan;49(1):13-26 [FREE Full text] [doi: 10.1016/j.jpainsymman.2014.05.003] [Medline: 24880002]
31. DuBenske LL, Gustafson DH, Shaw BR, Cleary JF. Web-based cancer communication and decision making systems: connecting patients, caregivers, and clinicians for improved health outcomes. Med Decis Making 2010;30(6):732-744 [FREE Full text] [doi: 10.1177/0272989X10386382] [Medline: 21041539]
32. Dy SM, Roy J, Ott GE, McHale M, Kennedy C, Kutner JS, et al. Tell Us™: a Web-based tool for improving communication among patients, families, and providers in hospice and palliative care through systematic data specification, collection, and use. J Pain Symptom Manage 2011 Oct;42(4):526-534 [FREE Full text] [doi: 10.1016/j.jpainsymman.2010.12.006] [Medline: 21458214]
33. McCall K, Keen J, Farrer K, Maguire R, McCann L, Johnston B, et al. Perceptions of the use of a remote monitoring system in patients receiving palliative care at home. Int J Palliat Nurs 2008 Sep;14(9):426-431. [doi: 10.12968/ijpn.2008.14.9.31121] [Medline: 19060793]
34. Ruland CM, Andersen T, Jeneson A, Moore S, Grimsbø GH, Børøsund E, et al. Effects of an Internet support system to assist cancer patients in reducing symptom distress: a randomized controlled trial. Cancer Nurs 2013;36(1):6-17. [doi: 10.1097/NCC.0b013e31824d90d4] [Medline: 22495503]
35. Snyder CF, Jensen R, Courtin SO, Wu AW, Website for Outpatient QOL Assessment Research Network. PatientViewpoint: a website for patient-reported outcomes assessment. Qual Life Res 2009 Sep;18(7):793-800 [FREE Full text] [doi: 10.1007/s11136-009-9497-8] [Medline: 19544089]
36. Williams GC. Improving patients' health through supporting the autonomy of patients and providers. In: Deci EL, Ryan RM, editors. Handbook of Self-Determination. Rochester, NY, USA: University of Rochester Press; 2002:233-254.
37. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319-340. [doi: 10.2307/249008]

38. Goodsmith N, Moore EM, Siddiq H, Barceló SE, Ulloa-Flores E, Loera G, et al. Community-partnered development of a digital mental health resource website to support diverse communities during the COVID-19 pandemic. J Health Care Poor Underserved 2022;33(1):506-516. [doi: 10.1353/hpu.2022.0039] [Medline: 35153238]

Abbreviations

CPPR: community-partnered participatory research
ESAS: Edmonton Symptom Assessment Scale
PRO: patient-reported outcome
QI: quality improvement

Edited by A Mavragani; submitted 26.09.21; peer-reviewed by S Manthri, R Halkes, A Carr; comments to author 19.01.22; revised version received 28.01.22; accepted 06.05.22; published 23.06.22.

Please cite as:
Al-Mondhiry J, D'Ambruoso S, Pietras C, Strouse T, Benzeevi D, Arevian AC, Wells KB
Co-created Mobile Apps for Palliative Care Using Community-Partnered Participatory Research: Development and Usability Study
JMIR Form Res 2022;6(6):e33849
URL: https://formative.jmir.org/2022/6/e33849
doi:10.2196/33849
PMID:35737441

Original Paper

Strategies and Lessons Learned During Cleaning of Data From Research Panel Participants: Cross-sectional Web-Based Health Behavior Survey Study

Mariana Arevalo[1], PhD; Naomi C Brownstein[2,3,4], PhD; Junmin Whiting[2], PhD; Cathy D Meade[1,3,5], RN, PhD; Clement K Gwede[1,3,5,6], RN, MPH, PhD; Susan T Vadaparampil[1,3,7], MPH, PhD; Kristin J Tillery[8], BA; Jessica Y Islam[3,7,9], PhD; Anna R Giuliano[3,7,9], PhD; Shannon M Christy[1,3,6,7], PhD

[1]Department of Health Outcomes and Behavior, Moffitt Cancer Center, Tampa, FL, United States
[2]Department of Biostatistics and Bioinformatics, Moffitt Cancer Center, Tampa, FL, United States
[3]Department of Oncological Sciences, University of South Florida, Tampa, FL, United States
[4]Department of Public Health Sciences, Medical University of South Carolina, Charleston, SC, United States
[5]Department of Genitourinary Oncology, Moffitt Cancer Center, Tampa, FL, United States
[6]Department of Gastrointestinal Oncology, Moffitt Cancer Center, Tampa, FL, United States
[7]Center for Immunization and Infection Research in Cancer, Moffitt Cancer Center, Tampa, FL, United States
[8]Participant Research, Interventions, and Measurement Core, Moffitt Cancer Center, Tampa, FL, United States
[9]Department of Cancer Epidemiology, Moffitt Cancer Center, Tampa, FL, United States

Corresponding Author:
Shannon M Christy, PhD
Department of Health Outcomes and Behavior
Moffitt Cancer Center
12902 Magnolia Drive
MFC-EDU
Tampa, FL, 33612
United States
Phone: 1 813 745 8840
Email: shannon.christy@moffitt.org

Abstract

Background: The use of web-based methods to collect population-based health behavior data has burgeoned over the past two decades. Researchers have used web-based platforms and research panels to study a myriad of topics. Data cleaning prior to statistical analysis of web-based survey data is an important step for data integrity. However, the data cleaning processes used by research teams are often not reported.

Objective: The objectives of this manuscript are to describe the use of a systematic approach to clean the data collected via a web-based platform from panelists and to share lessons learned with other research teams to promote high-quality data cleaning process improvements.

Methods: Data for this web-based survey study were collected from a research panel that is available for scientific and marketing research. Participants (N=4000) were panelists recruited either directly or through verified partners of the research panel, were aged 18 to 45 years, were living in the United States, had proficiency in the English language, and had access to the internet. Eligible participants completed a health behavior survey via Qualtrics. Informed by recommendations from the literature, our interdisciplinary research team developed and implemented a systematic and sequential plan to inform data cleaning processes. This included the following: (1) reviewing survey completion speed, (2) identifying consecutive responses, (3) identifying cases with contradictory responses, and (4) assessing the quality of open-ended responses. Implementation of these strategies is described in detail, and the Checklist for E-Survey Data Integrity is offered as a tool for other investigators.

Results: Data cleaning procedures resulted in the removal of 1278 out of 4000 (31.95%) response records, which failed one or more data quality checks. First, approximately one-sixth of records (n=648, 16.20%) were removed because respondents completed the survey unrealistically quickly (ie, <10 minutes). Next, 7.30% (n=292) of records were removed because they contained evidence of consecutive responses. A total of 4.68% (n=187) of records were subsequently removed due to instances of conflicting

responses. Finally, a total of 3.78% (n=151) of records were removed due to poor-quality open-ended responses. Thus, after these data cleaning steps, the final sample contained 2722 responses, representing 68.05% of the original sample.

Conclusions: Examining data integrity and promoting transparency of data cleaning reporting is imperative for web-based survey research. Ensuring a high quality of data both prior to and following data collection is important. Our systematic approach helped eliminate records flagged as being of questionable quality. Data cleaning and management procedures should be reported more frequently, and systematic approaches should be adopted as standards of good practice in this type of research.

(JMIR Form Res 2022;6(6):e35797) doi:10.2196/35797

KEYWORDS

data cleaning; data management; data integrity; quality assessment; research panel; web-based survey; interdisciplinary research; surveys and questionnaires; health behavior; internet

Introduction

The use of web-based methods to collect population-based data has burgeoned over the past two decades [1,2]. In fact, the number of published manuscripts reporting use of data from web-based platforms and research panels increased from 1 in 2010 to over 1200 in 2015 [3,4]. Research panels consist of individuals who volunteer to be contacted about potential participation in research studies [5,6]. Often, these research studies are available to potential participants via web-based platforms and may offer incentives for participation [5,6]. Researchers have used web-based platforms and research panels to study a wide variety of topics, including smoking cessation [7-9], social and behavioral determinants of health [10], eating habits [11], treatment seeking behaviors [12], social media use and experiences [13], participation in clinical trials research [14], virtual harassment and cyberbullying [15], addiction research [16,17], and infectious disease prevention behaviors [18,19], among others.

Web-based platforms and research panels are useful tools for recruiting and collecting information from large participant samples in a relatively short amount of time. More recently, these have become an alternative method for data collection due to COVID-19 pandemic restrictions (eg, social distancing). For example, after the start of the COVID-19 pandemic, one company with a research panel that allows researchers to reach potential participants via a web-based platform reported a 400% increase in the number of researchers using their platform [20]. Advantages to using these web-based platforms and research panels include the ability to assess a variety of behaviors, a high degree of diversity among potential participants, potentially lower research coordination costs, decreased time in data collection, and the ability to reach populations that otherwise would be difficult to recruit [6]. The number of users in these platforms and research panels have also increased, in part, due to the availability and ease of participation in research, the need to find supplementary income, or simply, for monetary gain [20]. However, fraudulent responses resulting from careless answering and the use of virtual personal networks to mask identities have contributed to a decline in data quality and integrity [21-23].

Only a few researchers have published their recommendations to improve the integrity of web-based survey data, and a combination of different strategies is advised [24-26]. Data integrity can be defined as the expectation of quality that is satisfactory and suitable to answer a research question [27]. In 2004, the Checklist for Reporting Results of Internet E-Surveys (CHERRIES) was published as a recommendation to ensure adequate reporting of web-based surveys [28]. One of the CHERRIES guidelines encourages researchers to report prevention methods for multiple survey entries from the same respondent, such as checking duplicate IP addresses and use of cookies. Although CHERRIES is helpful for improving researchers' reporting of findings from web-based survey studies, other data cleaning strategies to assess data quality are not specified in the guidelines.

In this paper, our study team shares our experiences in data cleaning to improve data quality and integrity from a web-based survey that recruited participants via a research panel. Our interdisciplinary team used a systematic and detailed data cleaning approach prior to the analyses. The goal for this paper is to describe our team's process and to share lessons learned, including a checklist developed by the team that other research teams could use or adapt to guide their data cleaning process.

Methods

Overview

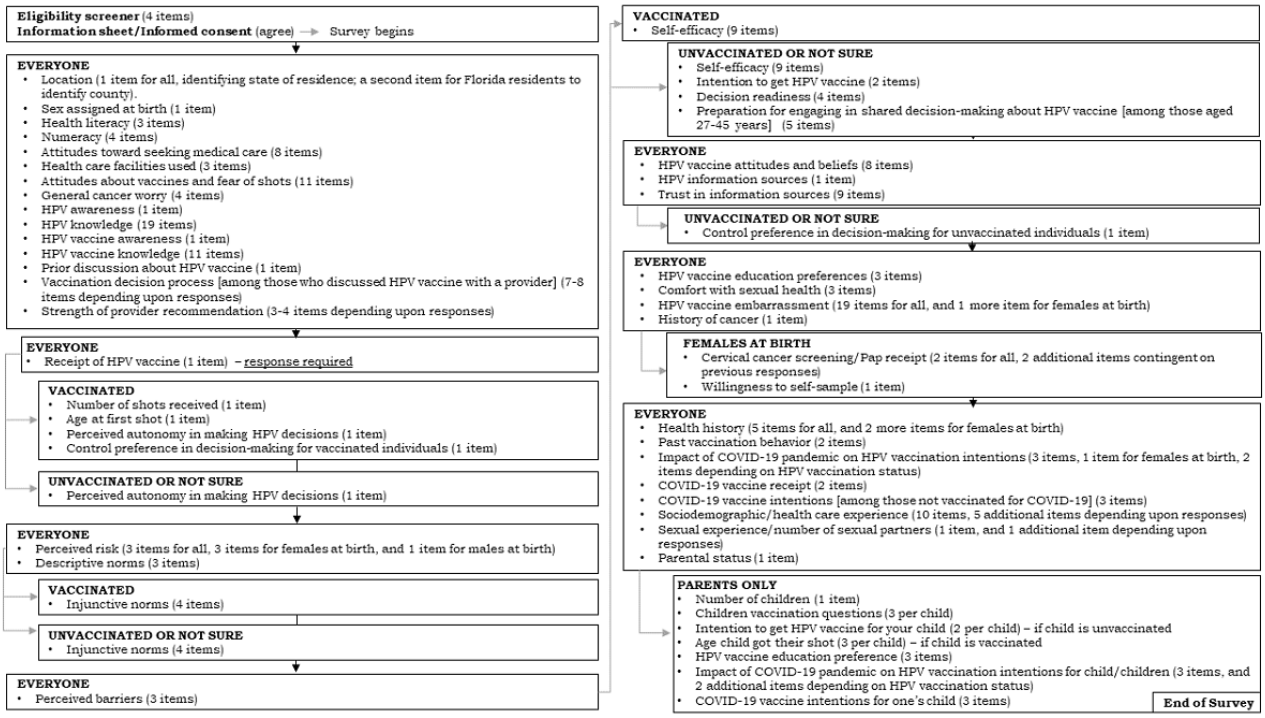
Data for this study were collected from panelists, either directly or via verified partners, of a research panel available for scientific and marketing research. The goal of our web-based survey was to examine human papillomavirus (HPV) and HPV vaccine knowledge, beliefs, attitudes, health care experiences, vaccine uptake, vaccination intentions, and other health behavior constructs, as well as information sources, preparedness for shared decision-making, and preferences that could help inform future HPV vaccination educational interventions for age-eligible individuals. Our interdisciplinary research team was composed of individuals with academic training in biostatistics, public health, nursing, psychology, epidemiology, and behavioral oncology.

Recruitment occurred from February 25 to March 24, 2021. The target sample for the study was 4000 individuals aged 18 to 45 years, stratified with equal recruitment by the cross-tabulation of age (18-26 years vs 27-45 years) and sex at birth (male vs female). Participation was limited to individuals who were panelists, directly or through verified partners of the research panel; were aged 18 to 45 years; were living in the United States; were proficient in the English language; and had access to the internet. Our interdisciplinary team aimed to recruit a sample

that was representative of the geographic as well as racial and ethnic characteristics in the United States. Florida residents and those within our cancer center's catchment were oversampled (ie, 500 Florida residents and 3500 residents of other states) with the aim of informing future research and outreach activities. The survey was pretested by three individuals who completed the paper-based version of the survey to estimate the completion time and provide feedback on survey item wording. Based on pretesting, it was estimated that the survey would take approximately 30 minutes to complete, depending on the sequential flow of the survey (ie, skip and contingency question patterns for some individuals).

The one-time survey was programmed in Qualtrics XM [29] by a member of the study team (KJT). The survey programmer applied Qualtrics features to monitor and set quota limits for gender and age counts. The final survey contained over 200 items. The number of items displayed for each respondent depended on the survey's branching logic, which was based on characteristics such as the respondent's age, sex assigned at birth, HPV vaccination status, and parental status (Figure 1). For example, the programmer set the branching logic such that respondents who self-identify as parents would receive a subset of questions regarding their children's health care experiences. Based on information that the panel company had about age, gender, and geographic location, potential participants were sent an invitation to participate in the study with a link to the survey directly by the research panel company. Individuals who were interested in participating completed a brief eligibility screener. Those who were eligible reviewed an informational sheet (ie, informed consent); eligible and interested individuals then proceeded to the main survey.

Figure 1. Survey schema illustrating branching logic, survey title, and number of items.



Ethical Considerations

The Scientific Review Committee at Moffitt Cancer Center and the Institutional Review Board of record (Advara) reviewed the study and approved it as exempt (Pro00047536).

Data Cleaning Strategies

Overview

Data quality is often defined in relation to aspects such as accuracy, completeness, validity, and conformity [30,31]. Evaluating the quality of web-based survey data and completing a data cleaning process before conducting statistical analyses is an important step that is often not reported transparently by research teams. To inform this process, our team conducted a literature review to identify key sources describing methods to support integrity of web-based survey responses. Findings from the literature review guided our team's decisions and helped us reach consensus on the number and types of strategies we would employ. Then, a systematic and sequential multi-strategy plan was developed to assess responses. The plan included the following: (1) reviewing survey completion speed [24,32], (2) identifying consecutive responses [32-34], (3) identifying cases with contradictory responses [35], and (4) assessing the quality of open-ended responses [26,35]. Thus, we defined high-quality data as survey data that had been stripped of instances of consecutive identical answers, contradictory responses, nonsensical open-ended responses, and responses completed in an unrealistic amount of time (see the steps below for further details on each criterion).

Step 1: Duration of Survey Completion

To identify a range of survey completion durations, a group of six individuals completed the survey. As previously mentioned, prior to survey launch, three individuals who were naïve to the survey items completed paper-and-pencil versions drafts of the survey to evaluate how long it might take potential participants

to complete the questions and provide feedback on item wording. Based upon the amount of time it took these individuals to complete the survey, our team anticipated that it would take participants an average of 30 minutes to complete the survey. However, we recognized that it could take some respondents less or more time to complete it depending on their responses and the corresponding skip logic that was programmed into the survey. For example, an individual who responded that they had received the HPV vaccine would receive items relevant to prior vaccine receipt, whereas an individual who reported that they had not received the HPV vaccine would receive questions about intentions to receive the HPV vaccine. Similarly, respondents with children would receive additional questions about HPV vaccination intentions for their children, whereas childless respondents would not receive those questions. Following finalization of the survey and the Qualtrics programming, an additional three team members completed the electronic (ie, Qualtrics) version of the survey.

Completion times of the Qualtrics-programmed survey within our team ranged from 5 minutes (when mindlessly and quickly clicking through the survey, but not actually reading the items) to 10 minutes (when reading and answering quickly, but legitimately) to 28 minutes (when attending to the items and reading thoroughly). Based on these test runs, consideration of the survey length and skip logic, and our best judgement, the team decided that a 10-minute (600-second) cutoff was the least amount of time that was still realistic in which respondents could take the survey while legitimately reading the items (see Table S1 in Multimedia Appendix 1 for descriptive information for Step 1).

Step 2: Consecutive Identical Responses

Consecutive identical responses (ie, "straight-lining" [34]) were assessed using four instruments that contained reverse-coded items and had been displayed to participants in table formats within the web-based survey. Selecting scales with reverse-coded items ensures that participant responses should not be identical in all items of a scale (see Table S1 in Multimedia Appendix 2 for a description of scales used in this step). Based on literature recommendations, we assessed consecutive identical responses of the response anchor extremes (ie, "strongly agree" and "strongly disagree") in the instrument's response scales [36,37]. For example, we identified records that demonstrated patterns of consecutive responses [34] on a vaccine attitudes scale, recognizing that individuals are unlikely to strongly agree or strongly disagree with both of the following statements: "Vaccines are generally safe" and "Vaccines are dangerous" [38]. A stepwise process was used to examine patterns of consecutive responses in four selected instruments: first, the instrument containing the largest number of items (ie, 19 items) was examined for patterns of consecutive responses, then survey records that failed this check were removed. These steps were repeated in the next scale, until all four scales had been checked. Records meeting those criteria were identified and removed using code in SAS software (version 9.4; SAS Institute Inc) [39] (Multimedia Appendix 3).

Step 3: Conflicting Responses

The team identified survey items that could indicate logical contradictions or extremely rare cases by carefully reviewing survey items and assessing patterns of responses for logical consistency. Depending on the survey content of a particular project, the types and numbers of questions used for assessment of conflicting answers might be different. For example, surveys might include the same question in two different locations of the survey (eg, age) to check for potential contradictory responses. During our data cleaning process, we decided to examine respondents who had indicated all of the following: (1) they were married or widowed or divorced, (2) they did not self-identify as asexual, (3) they reported that they had not ever had sexual intercourse (ie, vaginal, anal, or oral sex), and (4) they responded that they were a parent of one or more children. This group of cases was selected because we believe that it is an extremely unlikely scenario (ie, that one would be married or have a history of being married and have a child or children while reporting that they had never had sexual intercourse and did not self-identify as asexual) that is most likely due to careless answering. Records meeting those criteria were identified and removed using code in SAS software (version 9.4; SAS Institute Inc) [39] (Multimedia Appendix 4).

Step 4: Quality of Open-Ended Responses

Two team members independently assessed the quality of open-ended responses by checking all open-ended variables and identifying gibberish (ie, unintelligible responses), nonsensical responses (eg, responses that did not make sense in the context of the question asked), and patterns of identical responses within and across records (eg, exact same response to multiple open-ended items). Our team completed this in two steps. The first reviewer conducted a visual examination of cases that contained gibberish and duplicate responses. These records were flagged and removed. The second reviewer did the following: (1) identified nonsensical responses, (2) identified irrelevant responses, and (3) checked for repetitive patterns within and across records (ie, to identify whether different records had the same response patterns, because this could indicate that the same person may have completed multiple surveys). To do this, a team member (ie, first reviewer) exported survey records from SAS to a Microsoft Excel spreadsheet. Another team member (ie, second reviewer) located each open-ended variable column and sorted that column to inspect each of the responses provided, row by row. The team member also inspected the records column by column to identify patterns of open-ended responses across variables. Records that met the criteria outlined above were flagged. The same procedure was repeated for each open-ended variable until all open-ended variables in the codebook had been inspected. When the second reviewer had questions about whether or not responses were nonsensical, irrelevant, or repetitive, the research team discussed and resolved them by consensus. Survey records with instances of at least one of those three checks were flagged and subsequently removed from the data set.

Results

Table 1 describes the systematic and sequential steps leading to the final analytic sample consisting of 2722 records. About one-sixth of 4000 records (n=648, 16.20%) were flagged and removed during Step 1 (ie, survey completion duration). Descriptive statistics (ie, mean, median, and first and third quantiles) on completion duration for both the initial and final analytic samples are included in Multimedia Appendix 1. In Step 2, 7.30% (292/4000) of the records were removed because they contained evidence of consecutive responses. The SAS code for this step is included in Multimedia Appendix 3. In Step 3, 187 out of 4000 (4.68%) records were removed because we found evidence of conflicting responses. The SAS code for this step is included in Multimedia Appendix 4. In Step 4, 151 out of 4000 (3.78%) records were removed due to evidence of poor-quality open-ended responses. This final step required the most person-time effort, as some variables took up to 25 minutes to inspect. Ultimately, based on these four steps, 31.95% (1278/4000) of the responses from the original sample were removed.

We conducted descriptive statistics to characterize our sample before and after the quality assessment procedures (Table 2). The original sample was formed with equal groups based on age (18-26 years: 50%; 27-45 years: 50%) and sex at birth (females: 50%; males: 50%). After data cleaning, the final analytic sample (N=2722) contained a slightly higher number of females (55.95%) and individuals aged 18 to 26 years (50.73%). Compared to the original sample, the final sample had comparable proportions of individuals born in the United States and across sexual orientation categories. Also, in the final sample, a slightly higher proportion of respondents reported being White (68.80% original vs 71.05% final sample), non-Hispanic (81.30% original vs 83.25% final sample), childless (53.83% original vs 58.63% final sample), and from the Midwest region of the United States (20.28% original vs 21.42% final sample). Compared to the original sample, we observed a slightly lower proportion in the final sample of respondents with a graduate degree (21.20% original vs 15.76% final sample), with an annual income of US $100,000 or more (26.73% original vs 24.17% final sample), who were married (53.95% original vs 51.54% final sample), who were employed (74.33% original vs 72.56% final sample), and without health insurance (18.23% original vs 16.79% final sample). Manuscripts have been published [40], are under review, or are in preparation describing findings from this study.

Table 1. Steps to ensure quality of responses leading to final analytic sample.

Data quality assessment steps	All records (N=4000)	
	Records removed, n (%)	Records left, n (%)
Original sample	0 (0)	4000 (100)
Step 1: survey duration	648 (16.20)	3352 (83.80)
Step 2: consecutive identical responses	292 (7.30)	3060 (76.50)
Step 3: contradictory responses	187 (4.68)	2873 (71.83)
Step 4: quality of open-ended responses	151 (3.78)	2722 (68.05)

Table 2. Descriptive characteristics of the original and final samples.

Characteristic	Original sample (N=4000), n (%)[a]	Final sample (N=2722), n (%)[a]
Age (years)		
18-26	2000 (50.00)	1381 (50.73)
27-45	2000 (50.00)	1341 (49.27)
Sex assigned at birth		
Female	2000 (50.00)	1523 (55.95)
Male	2000 (50.00)	1199 (44.05)
Race		
White	2752 (68.80)	1934 (71.05)
Black or African American	506 (12.65)	314 (11.54)
Other	726 (18.15)	470 (17.27)
Missing	16 (0.40)	4 (0.15)
Ethnicity		
Hispanic	719 (17.98)	447 (16.42)
Non-Hispanic	3252 (81.30)	2266 (83.25)
Missing	29 (0.73)	9 (0.33)
Born in the United States		
Yes	3719 (92.98)	2529 (92.91)
No	263 (6.58)	189 (6.94)
Missing	18 (0.45)	4 (0.15)
Education		
High school or less	983 (24.58)	661 (24.28)
Some college or associate’s degree	1152 (28.80)	870 (31.96)
Bachelor’s degree	1000 (25.00)	757 (27.81)
Graduate school	848 (21.20)	429 (15.76)
Missing	17 (0.43)	5 (0.18)
Annual Income (US $)		
0-19,999	521 (13.03)	331 (12.16)
20,000-49,999	917 (22.93)	673 (24.72)
50,000-74,999	765 (19.13)	558 (20.50)
75,000-99,999	649 (16.23)	456 (16.75)
≥100,000 or more	1069 (26.73)	658 (24.17)
Missing	79 (1.98)	46 (1.69)
Relationship status		
Married	2158 (53.95)	1403 (51.54)
Other	1826 (45.65)	1317 (48.38)
Missing	16 (0.40)	2 (0.07)
Employment status		
Employed	2973 (74.33)	1975 (72.56)
Unemployed	415 (10.38)	310 (11.39)
Other	596 (14.90)	433 (15.91)
Missing	16 (0.40)	4 (0.15)
Sexual orientation		

Characteristic	Original sample (N=4000), n (%)[a]	Final sample (N=2722), n (%)[a]
Straight	3242 (81.05)	2225 (81.74)
Other	654 (16.35)	441 (16.20)
Missing	104 (2.60)	56 (2.06)
Health insurance status		
No	729 (18.23)	457 (16.79)
Yes	3248 (81.20)	2259 (82.99)
Missing	23 (0.58)	6 (0.22)
Parent to ≥1 child		
No	2153 (53.83)	1596 (58.63)
Yes	1832 (45.80)	1123 (41.26)
Missing	15 (0.38)	3 (0.11)
US geographic region		
Midwest	811 (20.28)	583 (21.42)
Northeast	680 (17.00)	435 (15.98)
South	1576 (39.40)	1072 (39.38)
West	933 (23.33)	632 (23.22)

[a]Percentages may not total 100% due to rounding.

Discussion

We described the use and systematic application of four steps to examine the quality of responses and to clean data from a web-based survey completed by individuals who were part of a research panel, directly or through verified partners. There are several other strategies and techniques to screen and clean data (eg, missing data, stability of response patterns, outliers, and maximum long strings, among others [24,32,33,41]). The types and number of strategies to use for data quality assessment and data cleaning may vary depending on the content, length, and complexity of the web-based survey (eg, access to IP addresses, attention checks, speeder flags, and naivety of respondents, among others [24,25,41,42]). For example, we used straight-lining to assess consecutive identical responses, and we selected scales that contained reverse-coded items to conduct this step. Alternatively, investigators who do not have scales containing reverse-coded items might consider other methods to assess consecutive responses, such as long-string analysis [33] or maximum long-string assessment [32]. Another note for other researchers to consider is that we did not assess respondents' IP addresses because we did not collect those data. There are benefits to collecting and examining IP addresses, such as identifying whether respondents took the survey more than once, but there may also be risks to collecting IP addresses, such as data protection and identity issues. Thus, investigators should consider risks and benefits when collecting IP addresses in their web-based surveys [25,43].

Other researchers have faced similar challenges when having to screen and filter out records with low-quality data collected from web-based surveys. Recently, researchers have reported disposing as much as three-quarters of data [44] or even their entire sample because over 90% of it was contaminated with fraudulent responses [35]. We lost about one-third of the original sample based on the criteria we used to clean the data. We recognize that losing this amount of data would be detrimental to an experimental design, but our study was observational and the final sample was sufficient for conducting our primary and exploratory analyses.

We hope that our step-by-step process encourages other research teams to systematically evaluate the integrity of their web-based survey data and use approaches to appropriately manage their data. Certainly, with the increased use of web-based surveys, it is imperative to evaluate data integrity and promote reporting transparency. A recent systematic review (n=80 studies) found that only 5% of the reviewed, published, web-based survey studies reported implementing checks to identify fraudulent data [45]. It is important to note that many panel companies may take steps to initially help ensure that panelists are participating in good faith (eg, not using bots to complete surveys) by using human intelligence tasks [3,4]. Many companies with research panels collect sociodemographic information about potential panelists and can send targeted study recruitment invitations based on the information initially reported to the company. As suggested by Dennis and colleagues [23], multiple entities have a responsibility and role in ensuring the integrity of the data [21,22].

This paper adds to the literature an applied, systematic example of data screening and management procedures that allow investigators to assess the quality of responses and eliminate invalid, fraudulent, or low-quality records. With the growing body of literature describing the application of quality assessment techniques and data cleaning approaches, this paper contributes an empirical example that could serve as a resource for other investigators and help streamline their data cleaning

procedures. We have created a checklist as a tool for future studies (Textbox 1).

Cleaning the data from our web-based survey completed by panelists was a multistep and time-consuming process. However, after having invested time and effort into these quality assessment and data cleaning steps, we are more confident about the integrity of the remaining data in our final analytic sample. The final sample for manuscripts resulting from this survey data may vary depending on scientific goals and data analysis decisions (ie, handling of missingness, among others).

There were several key lessons learned from this experience. First, screening and quality checks should be in place both before and after collecting data from web-based survey platforms and research panelists. In future web-based surveys, our team plans to include attention checks and additional items to assess conflicting answers with the hope of both decreasing and identifying the number of responses that are careless, fraudulent, or both. An example of an attention check is one published by Chandler and colleagues [41], in which participants were asked to select "satisfied" from a list of response options. This item, or other similar items, could help to flag inattentive respondents and those providing invalid data. Another key lesson learned was that web-based survey programming requires extensive attention to detail. We recommend that other investigators consider the steps outlined in our checklist (Textbox 1) for development and testing of their survey. Furthermore, we recommend that investigators consider the features and the constraints of the software package where their online survey will be programmed. For example, software formatting, features, and functions may limit the way in which items can be displayed and, therefore, how participants interact with these items. Thus, we recommend that other research teams both understand the functions and capabilities of the survey package to be used and conduct usability tests with a small number of respondents who can pilot-test the web-based survey prior to its launch. Ideally, the test takers should be from the target population and not be part of the research team to avoid familiarity with the survey. This will allow the researchers to identify any components of the survey display that might be unclear or confusing to participants and allow for an opportunity to reformat or change the survey items prior to the survey launch. Lastly, we learned that there are multiple ways to apply data quality checks, including removing records that follow a pattern or a series of flags, removing records with multiple flags in a sequential way, and using a single flag to remove records. There are trade-offs to each of these. For example, removing participants with any one of a number of possible flags is likely to decrease the number of careless and poor-quality responses, thereby increasing the data quality while also decreasing the sample size and, thus, power for further analyses. On the other hand, removing only participants who show evidence of poor quality by all flags decreases the chance of wrongly removing respondents who took the survey seriously but had one or more flag, such as speed reading, at the cost of leaving in respondents who may have poor quality data by some but not all criteria. Ultimately, our team decided that using multiple types of flags in a sequential order was an efficient way to identify and remove records with invalid data. Additionally, in keeping with good reporting practices, we recommend that investigators use the CHERRIES checklist [28] to ensure that information reported in their manuscripts follow recommended reporting guidelines, such as the following: descriptions of their study design, survey development and pretesting, recruitment process, survey administration details, response rates, prevention of multiple survey entries, and data analysis procedures relevant to electronic surveys.

Web-based survey data collection and the use of research panels will likely continue to be used by research teams in the future. Certainly, there are pros and cons to collecting web-based survey data and recruiting participants from research panels. Developing a rigorous plan throughout the study, from survey inception and survey development to survey administration and statistical analyses; using multiple strategies for data quality checks and cleaning; and devoting time and attention can be effective components of improving data cleaning and management practice and consequently increasing the integrity of web-based survey data.

Textbox 1. The Checklist for E-Survey Data Integrity.

Steps to develop and pretest an electronic survey:

- Provide clear instructions to participants and survey programmers
- Test skip and branching logic (ie, rules to jump to other items)
- Display items in a simple and logical way
- Display scales as individual items rather than in a table format
- Reduce the number of open-ended questions
- Reduce the use of complex fill-in tables
- Pretest the electronic survey in its final format for ease of administration and understanding (ideally with target population)
- Pretest the electronic survey for completion time
- Other (ie, other ways to tailor this checklist depending on needs and availability of data): _____

Steps to prevent fraudulent responses (pre–data collection):

- Add attention checks (ie, ways to identify inattentive respondents)
- Add CAPTCHA or reCAPTCHA tasks
- Add speeder checks (ie, ways to identify fast respondents)
- Add items that can be used to verify responses or assess contradictions
- Collect IP address, geolocation, device, and browser used to access the survey
- Enable settings available within the web-based survey application to prevent multiple submissions and detect bots, among other issues
- Choose a platform that adheres to data privacy and compliance
- Other (ie, other ways to tailor this checklist depending on needs and availability of data): _____

Steps to assess data integrity (post–data collection):

- Assess participant survey duration
- Check ranges of variables and examine responses that are clearly implausible
- Identify consecutive identical responses
- Identify contradictory responses
- Examine quality of open-ended responses
- Check IP address, geolocation, device, and browser information to identify multiple entries
- Other (ie, other ways to tailor this checklist depending on needs and availability of data): _____

Acknowledgments

This study was supported with funding from a Moffitt Center for Immunization and Infection Research in Cancer Award (principal investigator [PI]: SMC) and a Moffitt Merit Society Award (PI: SMC). This work has been supported, in part, by both the Participant Research, Interventions, and Measurement Core and the Biostatistics and Bioinformatics Shared Resource at the H. Lee Moffitt Cancer Center & Research Institute, a comprehensive cancer center designated by the National Cancer Institute and funded, in part, by a Moffitt Cancer Center Support Grant (P30-CA076292). NCB also acknowledges support from the South Carolina Clinical and Translational Science (SCTR) Institute at the Medical University of South Carolina. The SCTR Institute is funded by the National Center for Advancing Translational Sciences of the National Institutes of Health (grant UL1TR001450). The contents of this manuscript are solely the responsibility of the authors and do not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health.

The research team wishes to thank the following individuals: Katharine Head and Monica Kasting for their contribution to survey design, and Lakeshia Cousin, Aldenise Ewing, and Kayla McLymont for testing and timing early versions of the study survey.

Authors' Contributions

SMC designed the study and obtained funding. JW performed data analyses. SMC, NCB, JW, and MA conceived the idea for this manuscript. MA wrote the manuscript with input from all authors. NCB, JW, CDM, CKG, STV, KJT, JYI, ARG, and SMC critically reviewed and contributed to the final version of this manuscript.

Conflicts of Interest

NCB served as an ad hoc reviewer in 2020 for the American Cancer Society, for which she received sponsored travel during the review meeting and a stipend of US $300. NCB received a series of small awards for conference and travel support, including US $500 from the Statistical Consulting Section of the American Statistical Association (ASA) for Best Paper Award at the 2019 Joint Statistical Meetings and the US $500 Lee Travel Award from the Caucus for Women in Statistics to support attendance at the 2018 Joint Statistical Meetings. NCB also received a Michael Kutner/ASA Junior Faculty Travel Award of US $946.60 to attend the 2018 Summer Research Conference of the Southern Regional Council on Statistics and travel support of US $708.51 plus a registration waiver from the ASA to attend and chair a session for the 2017 Symposium on Statistical Inference. Currently, NCB serves as the Vice President for the Florida Chapter of the ASA and Section Representative for the ASA Statistical Consulting Section, and on the Scientific Review Board at Moffitt Cancer Center. Previously, NCB served as the Florida ASA Chapter Representative and as the mentoring subcommittee chair for the Regional Advisory Board of the Eastern North American Region of the International Biometrics Society. JYI has received consulting fees from Flatiron Health Inc. ARG is a member of Merck & Co, Inc, advisory boards and her institution receives funds from Merck & Co, Inc, for research. SMC is an unpaid advisory board member of HPV Cancers Alliance.

Multimedia Appendix 1
Descriptive information for Step 1—survey duration.
[DOCX File , 15 KB - formative_v6i6e35797_app1.docx]

Multimedia Appendix 2
Scales used to examine consecutive identical responses (Step 2).
[DOCX File , 16 KB - formative_v6i6e35797_app2.docx]

Multimedia Appendix 3
SAS code for Step 2 (consecutive identical responses).
[DOCX File , 16 KB - formative_v6i6e35797_app3.docx]

Multimedia Appendix 4
SAS code for Step 3 (conflicting responses).
[DOCX File , 15 KB - formative_v6i6e35797_app4.docx]

References

1. Internet surveys. Pew Research Center. URL: https://www.pewresearch.org/politics/methodology/collecting-survey-data/internet-surveys/ [accessed 2021-12-15]
2. Eysenbach G, Diepgen TL. Epidemiological data can be gathered with world wide web. BMJ 1998 Jan 03;316(7124):72 [FREE Full text] [doi: 10.1136/bmj.316.7124.72] [Medline: 9451290]
3. Bohannon J. Mechanical Turk upends social sciences. Science 2016 Jul 10;352(6291):1263-1264. [doi: 10.1126/science.352.6291.1263] [Medline: 27284175]
4. Rouse SV. Reliability of MTurk data from masters and workers. J Individ Differ 2020 Jan;41(1):30-36. [doi: 10.1027/1614-0001/a000300]
5. Gritz AS. Recruitment for online access panels. Int J Mark Res 2018 Jan 30;46(4):411-425. [doi: 10.1177/147078530404600406]
6. Gosling SD, Mason W. Internet research in psychology. Annu Rev Psychol 2015 Jan 03;66:877-902. [doi: 10.1146/annurev-psych-010814-015321] [Medline: 25251483]
7. Kulak J, Kamper-DeMarco K, Kozlowski L. Measuring support for requiring reduced nicotine cigarettes: Issues with questions, answers, and respondents. Health Behav Policy Rev 2020;7(6):546-556. [doi: 10.14485/hbpr.7.6.4]
8. Brewer NT, Jeong M, Hall MG, Baig SA, Mendel JR, Lazard AJ, et al. Impact of e-cigarette health warnings on motivation to vape and smoke. Tob Control 2019 Jul 10;28:e64-e70 [FREE Full text] [doi: 10.1136/tobaccocontrol-2018-054878] [Medline: 31292169]
9. Wackowski OA, Jeong M. Comparison of a general and conditional measure of e-cigarette harm perceptions. Int J Environ Res Public Health 2020 Jul 17;17(14):5151 [FREE Full text] [doi: 10.3390/ijerph17145151] [Medline: 32708854]

10. Giuse NB, Koonce TY, Kusnoor SV, Prather AA, Gottlieb LM, Huang L, et al. Institute of Medicine measures of social and behavioral determinants of health: A feasibility study. Am J Prev Med 2017 Mar;52(2):199-206 [FREE Full text] [doi: 10.1016/j.amepre.2016.07.033] [Medline: 27659121]
11. Hall MG, Lazard AJ, Grummon AH, Mendel JR, Taillie LS. The impact of front-of-package claims, fruit images, and health warnings on consumers' perceptions of sugar-sweetened fruit drinks: Three randomized experiments. Prev Med 2020 Mar;132:105998 [FREE Full text] [doi: 10.1016/j.ypmed.2020.105998] [Medline: 31982477]
12. Amsalem D, Martin A. Reducing depression-related stigma and increasing treatment seeking among adolescents: Randomized controlled trial of a brief video intervention. J Child Psychol Psychiatry 2022 Feb;63(2):210-217. [doi: 10.1111/jcpp.13427] [Medline: 33821507]
13. Vaterlaus JM, Spruance LA, Patten EV. Social media behaviors and experiences during the COVID-19 pandemic: Associations with anxiety, depression, and stress. J Soc Behav Health Sci 2021 Sep 27;15(1):217-230 [FREE Full text] [doi: 10.5590/jsbhs.2021.15.1.15]
14. Vazquez J, Abdelrahman S, Byrne LM, Russell M, Harris P, Facelli JC. Using supervised machine learning classifiers to estimate likelihood of participating in clinical trials of a de-identified version of ResearchMatch. J Clin Transl Sci 2020 Oct 04;5(1):e42 [FREE Full text] [doi: 10.1017/cts.2020.535] [Medline: 33948264]
15. Rafiq RI, Hosseinmardi H, Mattson SA, Han R, Lv Q, Mishra S. Analysis and detection of labeled cyberbullying instances in Vine, a video-based social network. Soc Netw Anal Min 2016 Sep 29;6(1):88. [doi: 10.1007/s13278-016-0398-x]
16. Mishra S, Carleton RN. Use of online crowdsourcing platforms for gambling research. Int Gambl Stud 2017 Feb 06;17(1):125-143. [doi: 10.1080/14459795.2017.1284250]
17. Strickland JC, Stoops WW. The use of crowdsourcing in addiction science research: Amazon Mechanical Turk. Exp Clin Psychopharmacol 2019 Mar;27(1):1-18. [doi: 10.1037/pha0000235] [Medline: 30489114]
18. Harper CA, Satchell LP, Fido D, Latzman RD. Functional fear predicts public health compliance in the COVID-19 pandemic. Int J Ment Health Addict 2021;19(5):1875-1888 [FREE Full text] [doi: 10.1007/s11469-020-00281-5] [Medline: 32346359]
19. Kelly BJ, Southwell BG, McCormack LA, Bann CM, MacDonald PDM, Frasier AM, et al. Predictors of willingness to get a COVID-19 vaccine in the US. BMC Infect Dis 2021 May 12;21(1):338 [FREE Full text] [doi: 10.1186/s12879-021-06023-9] [Medline: 33845781]
20. Peer E, Rothschild D, Gordon A, Evernden Z, Ekaterina E. Data quality of platforms and panels for online behavioral research. Behav Res Methods 2021 Oct 29:1-20 [FREE Full text] [doi: 10.3758/s13428-021-01694-3] [Medline: 34590289]
21. Ward M, Meade AW. Applying social psychology to prevent careless responding during online surveys. Appl Psychol 2017 Nov 10;67(2):231-263. [doi: 10.1111/apps.12118]
22. Meade AW, Craig SB. Identifying careless responses in survey data. Psychol Methods 2012 Oct;17(3):437-455. [doi: 10.1037/a0028085] [Medline: 22506584]
23. Dennis SA, Goodson BM, Pearson CA. Online worker fraud and evolving threats to the integrity of MTurk data: A discussion of virtual private servers and the limitations of IP-based screening procedures. Behav Res Account 2020;32(1):119-134. [doi: 10.2308/bria-18-044]
24. Newman A, Bavik YL, Mount M, Shao B. Data collection via online platforms: Challenges and recommendations for future research. Appl Psychol 2021 Jan 15;70(3):1380-1402. [doi: 10.1111/apps.12302]
25. Bernerth JB, Aguinis H, Taylor EC. Detecting false identities: A solution to improve web-based surveys and research on leadership and health/well-being. J Occup Health Psychol 2021 Dec;26(6):564-581. [doi: 10.1037/ocp0000281] [Medline: 34292017]
26. Kennedy R, Clifford S, Burleigh T, Waggoner PD, Jewell R, Winter NJG. The shape of and solutions to the MTurk quality crisis. Political Sci Res Methods 2020 Apr 24;8(4):614-629. [doi: 10.1017/psrm.2020.6]
27. Sandhu RS. On five definitions of data integrity. In: Proceedings of the 7th IFIP WG11.3 Working Conference on Database Security VII. 1993 Presented at: The 7th IFIP WG11.3 Working Conference on Database Security VII; September 12-15, 1993; Lake Guntersville, AL p. 257-267. [doi: 10.5555/646113.758632]
28. Eysenbach G. Improving the quality of web surveys: The Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
29. Qualtrics. URL: https://www.qualtrics.com/ [accessed 2022-03-31]
30. Jugulum R. Competing With High Quality Data: Concepts, Tools, and Techniques for Building a Successful Approach to Data Quality. Hoboken, NJ: John Wiley & Sons, Inc; 2014.
31. Schmidt CO, Struckmann S, Enzenbach C, Reineke A, Stausberg J, Damerow S, et al. Facilitating harmonized data quality assessments. A data quality framework for observational health research data collections with software implementations in R. BMC Med Res Methodol 2021 Apr 02;21(1):63 [FREE Full text] [doi: 10.1186/s12874-021-01252-7] [Medline: 33810787]
32. Niessen ASM, Meijer RR, Tendeiro JN. Detecting careless respondents in web-based questionnaires: Which method to use? J Res Pers 2016 Aug;63:1-11. [doi: 10.1016/j.jrp.2016.04.010]
33. Dupuis M, Meier E, Cuneo F. Detecting computer-generated random responding in questionnaire-based data: A comparison of seven indices. Behav Res Methods 2019 Oct;51(5):2228-2237. [doi: 10.3758/s13428-018-1103-y] [Medline: 30091086]

34. Kim Y, Dykema J, Stevenson J, Black P, Moberg DP. Straightlining: Overview of measurement, comparison of indicators, and effects in mail–web mixed-mode surveys. Soc Sci Comput Rev 2018 Feb 20;37(2):214-233. [doi: 10.1177/0894439317752406]
35. Burnette CB, Luzier JL, Bennett BL, Weisenmuller CM, Kerr P, Martin S, et al. Concerns and recommendations for using Amazon MTurk for eating disorder research. Int J Eat Disord 2022 Feb;55(2):263-272. [doi: 10.1002/eat.23614] [Medline: 34562036]
36. Curran PG. Methods for the detection of carelessly invalid responses in survey data. J Exp Soc Psychol 2016 Sep;66:4-19. [doi: 10.1016/j.jesp.2015.07.006]
37. Johnson JA. Ascertaining the validity of individual protocols from web-based personality inventories. J Res Pers 2005 Feb;39(1):103-129. [doi: 10.1016/j.jrp.2004.09.009]
38. Zimet GD, Weiss TW, Rosenthal SL, Good MB, Vichnin MD. Reasons for non-vaccination against HPV and future vaccination intentions among 19-26-year-old women. BMC Womens Health 2010 Oct 01;10:27 [FREE Full text] [doi: 10.1186/1472-6874-10-27] [Medline: 20809965]
39. SAS® 9.4 — Today and Tomorrow. SAS Support. URL: https://support.sas.com/software/94/ [accessed 2022-05-31]
40. Brownstein NC, Reddy H, Whiting J, Kasting ML, Head KJ, Vadaparampil ST, et al. COVID-19 vaccine behaviors and intentions among a national sample of United States adults ages 18-45. Prev Med 2022 May 07;160:107038 [FREE Full text] [doi: 10.1016/j.ypmed.2022.107038] [Medline: 35398369]
41. Chandler J, Rosenzweig C, Moss AJ, Robinson J, Litman L. Online panels in social science research: Expanding sampling methods beyond Mechanical Turk. Behav Res Methods 2019 Oct;51(5):2022-2038 [FREE Full text] [doi: 10.3758/s13428-019-01273-7] [Medline: 31512174]
42. Robinson J, Rosenzweig C, Moss AJ, Litman L. Tapped out or barely tapped? Recommendations for how to harness the vast and largely unused potential of the Mechanical Turk participant pool. PLoS One 2019;14(12):e0226394 [FREE Full text] [doi: 10.1371/journal.pone.0226394] [Medline: 31841534]
43. Teitcher JEF, Bockting WO, Bauermeister JA, Hoefer CJ, Miner MH, Klitzman RL. Detecting, preventing, and responding to "fraudsters" in internet research: Ethics and tradeoffs. J Law Med Ethics 2015;43(1):116-133 [FREE Full text] [doi: 10.1111/jlme.12200] [Medline: 25846043]
44. Pratt-Chapman M, Moses J, Arem H. Strategies for the identification and prevention of survey fraud: Data analysis of a web-based survey. JMIR Cancer 2021 Jul 16;7(3):e30730 [FREE Full text] [doi: 10.2196/30730] [Medline: 34269685]
45. Sharma R, Tikka SK, Bhute AR, Bastia BK. Adherence of online surveys on mental health during the early part of the COVID-19 outbreak to standard reporting guidelines: A systematic review. Asian J Psychiatr 2021 Nov;65:102799 [FREE Full text] [doi: 10.1016/j.ajp.2021.102799] [Medline: 34365280]

Abbreviations

ASA: American Statistical Association
CHERRIES: Checklist for Reporting Results of Internet E-Surveys
HPV: human papillomavirus
PI: principal investigator
SCTR: South Carolina Clinical and Translational Science

Edited by A Mavragani; submitted 17.12.21; peer-reviewed by A Loban, T Karen, K Gierend; comments to author 09.03.22; revised version received 13.05.22; accepted 20.05.22; published 23.06.22.

Please cite as:
Arevalo M, Brownstein NC, Whiting J, Meade CD, Gwede CK, Vadaparampil ST, Tillery KJ, Islam JY, Giuliano AR, Christy SM Strategies and Lessons Learned During Cleaning of Data From Research Panel Participants: Cross-sectional Web-Based Health Behavior Survey Study
JMIR Form Res 2022;6(6):e35797
URL: https://formative.jmir.org/2022/6/e35797
doi:10.2196/35797
PMID:35737436

Original Paper

Mobile-Based and Self-Service Tool (iPed) to Collect, Manage, and Visualize Pedigree Data: Development Study

Chen Sun[1*], MA; Jing Xu[1*], PhD; Junxian Tao[1*], MA; Yu Dong[1], MA; Haiyan Chen[1], MA; Zhe Jia[1], MA; Yingnan Ma[1], MA; Mingming Zhang[1], PhD; Siyu Wei[1], MA; Guoping Tang[2], MA; Hongchao Lyu[1], MA; Yongshuai Jiang[1], PhD

[1]College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, China

[2]The Fourth Affiliated Hospital, Zhejiang University School of Medicine, Zhejiang, China

[*]these authors contributed equally

Corresponding Author:
Yongshuai Jiang, PhD
College of Bioinformatics Science and Technology
Harbin Medical University
194 Xuefu Road
Nangang District
Harbin, 150081
China
Phone: 86 451 86620941
Email: jiangyongshuai@gmail.com

Abstract

Background: Pedigree data (family history) are indispensable for genetics studies and the assessment of individuals' disease susceptibility. With the popularity of genetics testing, the collection of pedigree data is becoming more common. However, it can be time-consuming, laborious, and tedious for clinicians to investigate all pedigree data for each patient. A self-service robot could inquire about patients' family history in place of professional clinicians or genetic counselors.

Objective: The aim of this study was to develop a mobile-based and self-service tool to collect and visualize pedigree data, not only for professionals but also for those who know little about genetics.

Methods: There are 4 main aspects in the iPed construction, including interface building, data processing, data storage, and data visualization. The user interface was built using HTML, JavaScript libraries, and Cascading Style Sheets (version 3; Daniel Eden). Processing of the submitted data is carried out by PHP programming language. MySQL is used to document and manage the pedigree data. PHP calls the R script to accomplish the visualization.

Results: iPed is freely available to all users through the iPed website. No software is required to be installed, no pedigree files need to be prepared, and no knowledge of genetics or programs is required. The users can easily complete their pedigree data collection and visualization on their own and through a dialogue with iPed. Meanwhile, iPed provides a database that stores all users' information. Therefore, when the users need to construct new pedigree trees for other genetic traits or modify the pedigree trees that have already been created, unnecessary duplication of operations can be avoided.

Conclusions: iPed is a mobile-based and self-service tool that could be used by both professionals and nonprofessionals at any time and from any place. It reduces the amount of time required to collect, manage, and visualize pedigree data.

(JMIR Form Res 2022;6(6):e36914) doi:10.2196/36914

KEYWORDS

pedigree; pedigree data; visualization; self-service; mobile-based

Introduction

Genetics plays a vital role in all diseases, and therefore, pedigree data are frequently used for clinical diagnosis [1]. Over the past few years, more attention has been paid to formulating an efficient model for disease risk prediction. Accurate risk prediction can be of significant help in disease prevention and follow-up strategies [2]. There is consensus that family history is an important risk factor in estimating an individual's susceptibility to diseases, especially hereditary diseases, as it

reflects the shared heredity and environment [3-5]. Numerous studies have applied family history as a risk factor to study the treatment and outcomes of schizophrenia [6,7]; predict cardiovascular disease risk [8]; and assess the risk of common cancers such as gastric cancer [9], colorectal cancer [10], and breast cancer [11]. However, pedigree data collection is so limited by familiarity and family communication of patients that it is hard to acquire in a clinical environment [12]. In addition, genetic information will change over time as the number of family members and confirmed relatives increases [12].

To address these issues, several web-based tools have been developed to collect and visualize pedigree data. However, it is indispensable for clinicians and genetics counselors to possess professional knowledge in order to use these tools. For example, *Pedigreejs* [13] is a pedigree editor based on JavaScript that can produce SVG format images in web browsers [14]; however, it is not easy to use for many users who are unfamiliar with JavaScript or specific web programming visualization libraries [15]. *Ped_draw* [15] can generate an image file as a command line or web tool on the condition that users prepare pedigree data first. *Panogram* [16] and *HaploPainter* [17] are both software to visualize the pedigree data, while the image file is not free to use.

We developed iPed, a mobile-based and self-service tool, to collect and visualize pedigree data. Mobile-based technologies are more widespread than web-based technologies, and they can improve the quality of life and reduce health care costs effectively [18,19]. Collection and visualization of pedigree data is accomplished by responding to some questions posed by iPed. There is no need to prepare an input file or perform any other operation except tapping on some options on the phone. iPed's intelligent robot significantly reduces the work pressure of the clinician in querying about the patient's family history. Furthermore, by storing pedigree data in a database, iPed makes it more convenient for users to manage their information.

Methods

Overview of Modules

iPed has 3 main parts: (1) *Your pedigree*, which stores information about family members; (2) *History*, which includes the phenotypic information users have previously entered and the visualizations; (3) *New*, which collects family information about a selected or newly created phenotype through answers to some questions asked by iPed. The interface was built using HTML, which includes a set of tags that unify the format of documents on the network and connect scattered internet resources into a logical whole. JavaScript libraries are used to respond to browser events, including changing information of family members in *Your pedigree*, modifying the phenotype data in *History*, and the dialogue with iPed when creating a new phenotype in *New*. The combination of HTML and JavaScript can complete the interaction between users and iPed. In addition, Cascading Style Sheets (version 3; Daniel Eden), a language that defines style structures such as font, color, and position, is applied to modify web pages.

Data Processing

ThinkPHP (version 3.1.3; Chen Liu et al), a fast, compatible, and simple lightweight PHP development framework, is applied for the processing of the submitted data. The user could tap on the options to answer questions asked by iPed. When the dialogue ends, the data collection is finished at the same time. iPed will then convert the data into a specific format for further storage and visualization.

Data Storage

All data including family members' age and phenotypic information are stored in MySQL (version 5.0; Oracle Corporation), the most commonly used relational database management system. PhpMyAdmin (version 3.3.7; The phpMyAdmin Project) is a PHP and web-based MySQL database management tool for managing MySQL databases with a web interface.

Pedigree Data Visualization

iPed completes the data visualization by calling the R script in PHP. Kinship2 [20], an R package restructured from kinship package [21], is used to visualize the pedigree data. Conventionally, squares and circles represent males and females, respectively. If the person has the phenotype, the pattern will be colored black; otherwise, it will be white. The connecting lines between patterns indicate genetic relationships.

Ethics Consideration

This study did not require ethics approval as it did not involve any human subjects and the mobile-based tool was developed for visualization of pedigree data for all users.

Results

Pedigree and First Phenotype Data Collection

iPed is a mobile-based, web-based, self-service tool for collecting and visualizing pedigree data on genetic phenotypes. A first-time user needs to register to save personal information for the next use. After a successful login, the user can select a phenotype of interest (Figure 1B), including disease phenotypes (eg, lung cancer, breast cancer, and rheumatoid arthritis), common phenotypes (eg, double eyelid, tall, and right-handed), and entertaining phenotypes (eg, singing well, high income, or social phobias). The user can also create a new phenotype and complete the family data collection for the created phenotype through an interactive dialogue with an intelligent robot (Figure 1C). The robot will ask some questions about the user's age, whether the user has the phenotype, whether the user has a spouse or child, and whether the user's relatives and their families have the phenotype (Figure 2).

After the first use, iPed will save the user's family information in *YourPedigree* (Figure 1F), including family members and their ages. This makes it convenient to directly collect pedigree data about the user's family members when the user describes a new genetic phenotype.

When the users log in again, they can check their family information (*Your pedigree*), view the genetic phenotypic information (eg, the phenotype name, such as *dimple*), and the

corresponding visualization (*Visualize the pedigree*). They can also select a new phenotype on the home page (*New*) to visualize the information (Figure 1A).

Figure 1. Main pages when the user logs back into iPed the next time. (A) The home page; (B) alternative phenotypes on iPed when the user taps on *New*; (C) the dialogue between the user and the robot when the user needs to define a new phenotype; (D) visualization of the genetic phenotypic data; (E) the genetic phenotype the user has entered, including the family members and their phenotypic information; (F) *Your pedigree* shows the family members and their ages.

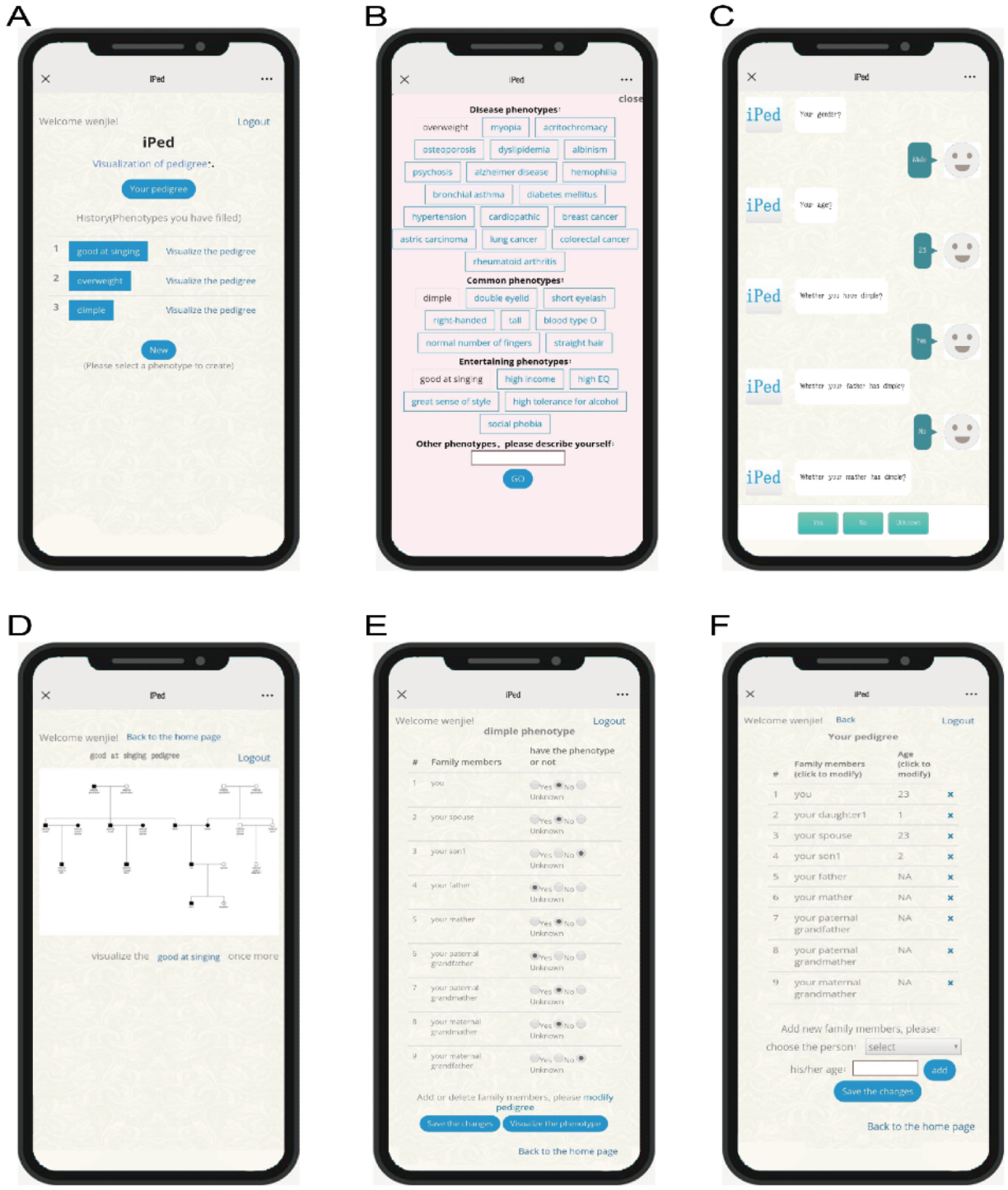
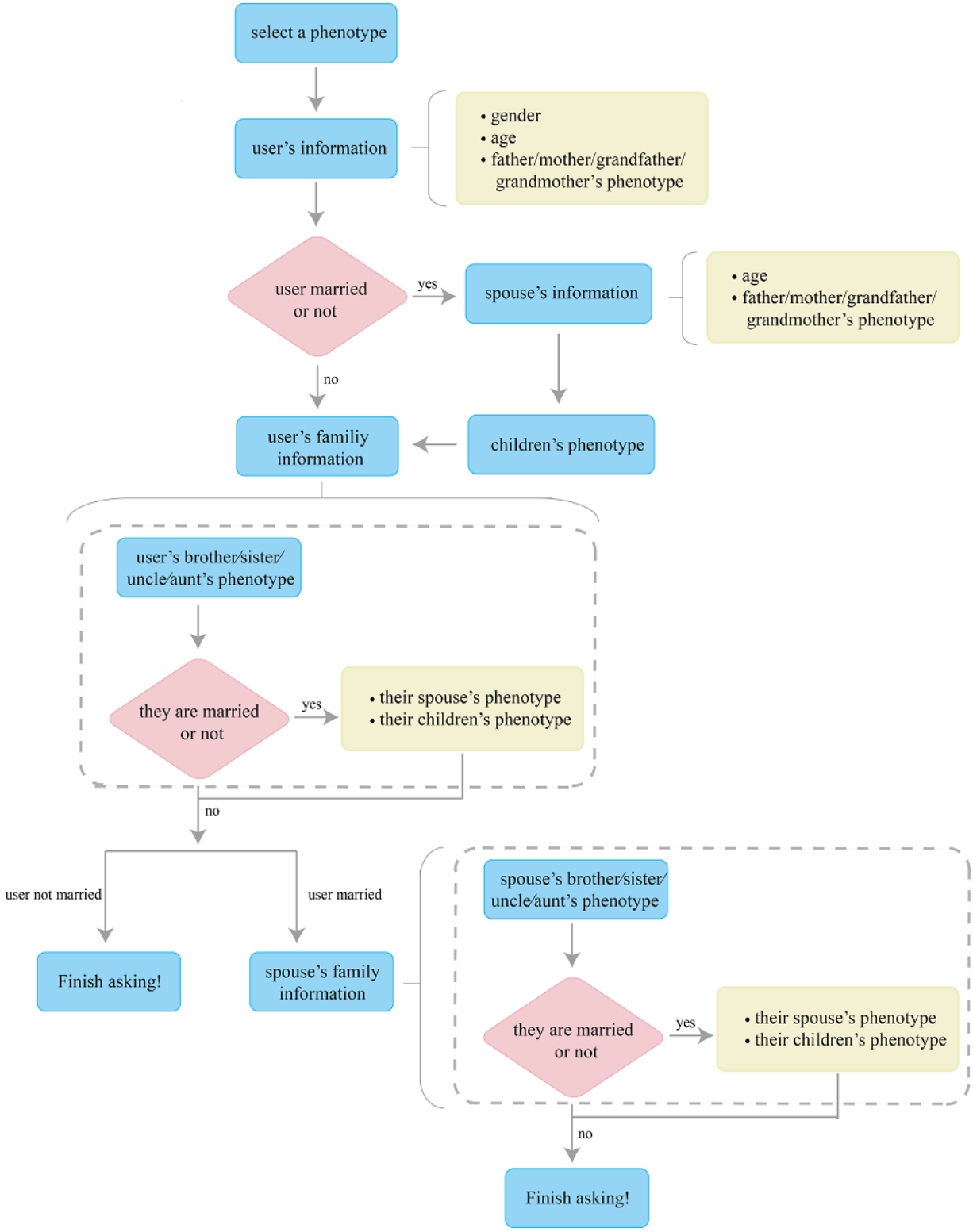


Figure 2. The flow path of the questions asked by the intelligent robot when the user logs into iPed for the first time.



Phenotype Data Visualization and Modification

Phenotype Data Visualization

Phenotype data visualization is accessible by tapping on *Visualize the pedigree*, and the user can see the visualized phenotypic information on the result page (Figure 1D). When the users find information that needs to be changed, they can tap on *Visualize the phenotype once more* to complete the modification.

Phenotype Data Modification

When information needs to be modified, the user can just click on the name of phenotype such as *dimple* (Figure 1A) or tap on *Visualize the phenotype once more* on the result page (Figure 1D) and change the targeted family member's status on the page showing phenotypic information (Figure 1E). If the family member's status is changed, tapping on *Modify pedigree* (Figure 1E) will direct the user to *Your pedigree*. *YourPedigree* provides options to add or delete family members and modify their ages.

It should be noted that the already existing family members cannot be added again. After modifying the family member's information, the user can return to change the member's phenotypic information. After finishing all the modifications, the user can visualize the phenotypic information afresh.

New Phenotype Data Adding

If users intend to visualize new phenotypic data, they could tap on *New* on the home page, select an unfulfilled phenotype that they are interested in (Figure 1B), and complete the dialogue with the robot. The robot will simply ask whether the members added by the user have this phenotype. Afterward, it will generate the corresponding visualization.

Discussion

Principal Findings

Pedigree data are crucial in genetics studies and disease diagnosis, development, and prognosis. There are thousands of studies that have applied family history as a risk factor. However, this information is not easy to collect and manage. iPed provides a self-service tool for users to collect and visualize this information. With iPed, users do not need to prepare an input file first or learn any knowledge about genetics. They can visualize their pedigree data just by answering the intelligent robot's questions, without any professional operations. At the same time, all the information submitted by users is stored in a database, so that visualizing a new phenotype or modifying previously entered data is more convenient.

Comparison With Prior Work

Recently, some web-based tools for collecting and visualizing pedigree data have been developed. There are some limitations to these existing tools; if the user would like to use these tools, they have to gain some knowledge first. For example, *Pedigreejs* [13] is a web-based tool that is difficult to use for those who have little knowledge about JavaScript or specific web programming visualization libraries. *HaploForge* [22] is another web application that is hard to use, as the meaning of the different symbols and connecting lines can be confusing for users who do not know about genetics; it is always difficult to know how to proceed to the next operation. *Ped_draw* [23] can be used only when users prepare pedigree data first. In addition, many tools such as *panogram* [16] and *HaploPainter* [17] are not free to use for all users. There are also some tools such as *PediDraw* [24] and *Madeline 2.0 PDE* [25] that were first released over a decade ago, and access to them is presently unstable or even disabled.

iPed is a mobile-based and self-service tool that can be used easily in place of software needing installation or web-based tools requiring professional knowledge. Collecting pedigree data is easier, as users only need to answer some questions asked by an intelligent robot and tap on some options. It is a simple and efficient tool for both professional and nonprofessional uses. Meanwhile, iPed provides a database to save all information submitted by users, providing greater convenience for subsequent uses. Additionally, iPed is free for everyone to use.

Strengths and Limitations

iPed enables the collection of pedigree data through dialogue with an intelligent robot. As the pedigree data are usually complex, the possibility of missing some information is high, and that can lead to wrong conclusions. The dialogue with iPed helps the user recall their pedigree data more comprehensively. Questions asked by iPed include the following: "How many uncles do you have?" "Is your uncle married?" "How many children does your uncle have?" and "Do your uncle's children have the phenotype?" Without any other operations, the users can answer the questions just by tapping on the "yes," "no," or "unknown" options, and the pedigree data collection is accomplished when the dialogue ends. The pedigree data visualization will then generate automatically. Moreover, considering that there are associations between diverse phenotypes, iPed will save all the information submitted by the users after the first use. Therefore, it is easy to look over the phenotypes that have been entered before. iPed offers substantial help in clinical diagnosis of complications and genetics studies about the correlations of different phenotypes.

Although iPed provides many novel functions, there are still some limitations. First, in the visualization of pedigree data, the color black traditionally indicates a person with the phenotype and white indicates one without the phenotype. iPed will offer a more powerful function if more colors are introduced, indicating different meanings. Second, users should select a phenotype before the pedigree data collection, so that the resultant picture shows information for the specified phenotype. How to present multiple phenotypes in a single picture will be considered in the future. We will constantly update and upgrade iPed to offer a significantly better user experience.

Conclusions

iPed was developed as a mobile-based and self-service tool to collect and visualize pedigree data [26]; it can be used by professional researchers, clinicians, and those who possess little relevant knowledge. iPed shortens the amount of time patients spend in the hospital and improves the efficiency of the clinicians. With iPed, collection, management, and use of pedigree data will no longer be difficult.

Acknowledgments

This work is funded by National Natural Science Foundation of China (grants 31970651, and 92046018). It was also supported by the Tianyuan Fund for Mathematics of the National Natural Science Foundation of China (grant 12026414), and Basic Public Welfare Research Project of Zhejiang Province (grant LGF20C060001).

Authors' Contributions

YJ conceptualized and designed the work. CS, JX, JT, HC, and YD performed research and drafted and modified the manuscript. All authors contributed to discussing and revising the manuscript.

Conflicts of Interest

None declared.

References

1. Jackson M, Marks L, May GHW, Wilson JB. The genetic basis of disease. Essays Biochem (2018) 2018 Dec;62(5):643-723. [doi: 10.1042/ebc20170053] [Medline: 30509934]
2. Jostins L, Barrett JC. Genetic risk prediction in complex disease. Hum Mol Genet 2011 Oct 15;20(R2):R182-R188 [FREE Full text] [doi: 10.1093/hmg/ddr378] [Medline: 21873261]
3. Ofori SN, Odia OJ. Risk assessment in the prevention of cardiovascular disease in low-resource settings. Indian Heart J 2016 May;68(3):391-398 [FREE Full text] [doi: 10.1016/j.ihj.2015.07.004] [Medline: 27316504]
4. Jacobs MF. Predicting cancer risk based on family history. Elife 2021 Sep 29;10:10 [FREE Full text] [doi: 10.7554/eLife.73380] [Medline: 34586069]
5. Igo RP, Kinzy TG, Cooke Bailey JN. Genetic risk scores. Curr Protoc Hum Genet 2019 Dec 25;104(1):e95 [FREE Full text] [doi: 10.1002/cphg.95] [Medline: 31765077]
6. Mortensen PB, Pedersen MG, Pedersen CB. Psychiatric family history and schizophrenia risk in Denmark: which mental disorders are relevant? Psychol Med 2009 Jul 17;40(2):201-210. [doi: 10.1017/s0033291709990419] [Medline: 19607751]
7. Käkelä J, Marttila R, Keskinen E, Veijola J, Isohanni M, Koivumaa-Honkanen H, et al. Association between family history of psychiatric disorders and long-term outcome in schizophrenia - the Northern Finland birth cohort 1966 study. Psychiatry Res 2017 Mar;249:16-22. [doi: 10.1016/j.psychres.2016.12.040] [Medline: 28063393]
8. Mehta A, Virani SS, Ayers CR, Sun W, Hoogeveen RC, Rohatgi A, et al. Lipoprotein(a) and family history predict cardiovascular disease risk. J Am Coll Cardiol 2020 Aug 18;76(7):781-793 [FREE Full text] [doi: 10.1016/j.jacc.2020.06.040] [Medline: 32792075]
9. Yuan G, Chen Y, He S. Family history of gastric cancer and helicobacter pylori treatment. N Engl J Med 2020 May 28;382(22):2171-2172. [doi: 10.1056/NEJMc2003542] [Medline: 32459942]
10. Pesola F, Eloranta S, Martling A, Saraste D, Smedby KE. Family history of colorectal cancer and survival: a Swedish population-based study. J Intern Med 2020 Jun 03;287(6):723-733 [FREE Full text] [doi: 10.1111/joim.13036] [Medline: 32012369]
11. Brewer HR, Jones ME, Schoemaker MJ, Ashworth A, Swerdlow AJ. Family history and risk of breast cancer: an analysis accounting for family structure. Breast Cancer Res Treat 2017 Aug 3;165(1):193-200 [FREE Full text] [doi: 10.1007/s10549-017-4325-2] [Medline: 28578505]
12. Helfand B. A comparison of genetic risk score with family history for estimating prostate cancer risk. Asian J Androl 2016;18(4):515. [doi: 10.4103/1008-682x.177122]
13. Pedigreejs. URL: https://ccge-boadicea.github.io/pedigreejs/ [accessed 2022-06-10]
14. Carver T, Cunningham A, Babb de Villiers C, Lee A, Hartley S, Tischkowitz M, et al. pedigreejs: a web-based graphical pedigree editor. Bioinformatics 2018 Mar 15;34(6):1069-1071 [FREE Full text] [doi: 10.1093/bioinformatics/btx705] [Medline: 29095980]
15. Velinder M, Lee D, Marth G. ped_draw: pedigree drawing with ease. BMC Bioinformatics 2020 Dec 09;21(1):569 [FREE Full text] [doi: 10.1186/s12859-020-03917-4] [Medline: 33297934]
16. panogram. URL: http://panogram.github.io/panogram/ [accessed 2022-06-10]
17. HaploPainter. SourceForge. URL: https://sourceforge.net/projects/haplopainter/ [accessed 2022-06-10]
18. Ayatollahi H, Ghalandar Abadi M, Hemmat M. Web and mobile-based technologies for monitoring high-risk pregnancies. BMJ Health Care Inform 2019 Jun 14;26(1):e000025 [FREE Full text] [doi: 10.1136/bmjhci-2019-000025] [Medline: 31201201]
19. Emami H, Asadi F, Garavand A. The features of mobile-based software in self-management of patients with asthma: a review article. Tanaffos 2020 Jan;19(1):10-19 [FREE Full text] [Medline: 33101427]
20. Sinnwell JP, Therneau TM, Schaid DJ. The kinship2 R package for pedigree data. Hum Hered 2014 Jul 29;78(2):91-93 [FREE Full text] [doi: 10.1159/000363105] [Medline: 25074474]
21. Zhao JH. Pedigree-drawing with R and graphviz. Bioinformatics 2006 Apr 15;22(8):1013-1014. [doi: 10.1093/bioinformatics/btl058] [Medline: 16488908]
22. HaploForge. URL: https://github.com/mtekman/HaploForge [accessed 2022-06-10]
23. Ped_draw. URL: https://peddraw.github.io [accessed 2022-06-10]
24. He M, Li W. PediDraw: a web-based tool for drawing a pedigree in genetic counseling. BMC Med Genet 2007 Jun 08;8(1):31 [FREE Full text] [doi: 10.1186/1471-2350-8-31] [Medline: 17555606]

25. Trager EH, Khanna R, Marrs A, Siden L, Branham KE, Swaroop A, et al. Madeline 2.0 PDE: a new program for local and web-based pedigree drawing. Bioinformatics 2007 Jul 15;23(14):1854-1856. [doi: 10.1093/bioinformatics/btm242] [Medline: 17488757]
26. iPed. URL: http://www.onethird-lab.com/iPed/ [accessed 2022-06-10]

Edited by A Mavragani; submitted 29.01.22; peer-reviewed by Q Zhang, UH Mohamad, CH Lai; comments to author 28.04.22; revised version received 15.05.22; accepted 06.06.22; published 23.06.22.

Please cite as:
Sun C, Xu J, Tao J, Dong Y, Chen H, Jia Z, Ma Y, Zhang M, Wei S, Tang G, Lyu H, Jiang Y
Mobile-Based and Self-Service Tool (iPed) to Collect, Manage, and Visualize Pedigree Data: Development Study
JMIR Form Res 2022;6(6):e36914
URL: https://formative.jmir.org/2022/6/e36914
doi:10.2196/36914
PMID:35737451

Original Paper

Development of a Head-Mounted Holographic Needle Guidance System for Enhanced Ultrasound-Guided Neuraxial Anesthesia: System Development and Observational Evaluation

Jaya Tanwani[1], MD; Fahad Alam[1], MD; Clyde Matava[2], MD; Stephen Choi[1], MD; Paul McHardy[1], MD; Oskar Singer[1], MD; Geraldine Cheong[3], MBBS; Julian Wiegelmann[1], MD

[1]Department of Anesthesiology and Pain Medicine, Sunnybrook Health Sciences Centre, University of Toronto, Toronto, ON, Canada

[2]Department of Anesthesia, The Hospital for Sick Children, University of Toronto, Toronto, ON, Canada

[3]Department of Anesthesia, Khoo Teck Puat Hospital, Singapore, Singapore

Corresponding Author:
Jaya Tanwani, MD
Department of Anesthesiology and Pain Medicine
Sunnybrook Health Sciences Centre
University of Toronto
2075 Bayview Avenue
Toronto, ON, M4N 3M5
Canada
Phone: 1 416 480 6100
Email: jayatanwani9@gmail.com

Abstract

Background: Neuraxial anesthesia is conventionally performed using a landmark-based technique. Preprocedural ultrasound is often used in challenging clinical scenarios to identify an ideal needle path. The procedure is then carried out by the operator recreating the ultrasound needle path from memory. We suggest that a needle guidance system using the Microsoft HoloLens mixed reality headset, which projects a hologram of the ideal needle path, can assist operators in replicating the correct needle angulation and result in fewer needle passes.

Objective: The objective of the study was to develop software for the mixed reality HoloLens headset, which could be used to augment the performance of neuraxial anesthesia, and establish its face validity in lumbar spine phantom models.

Methods: We developed an ultrasound transducer marker and software for the HoloLens, which registers the position and angulation of the ultrasound transducer during preprocedural scans. Once an image of a clear path from skin to the intrathecal space is acquired, a hologram of the ideal needle path is projected onto the user's visual field. The ultrasound probe is removed while the hologram remains in the correct spatial position to visualize the needle trajectory during the procedure as if conducting real-time ultrasound. User testing was performed using a lumbar spine phantom.

Results: Preliminary work demonstrates that novice (2 anesthesia residents) and experienced operators (5 attending anesthesiologists) can rapidly learn to use mixed reality holograms to perform neuraxial anesthesia on lumbar spine phantoms.

Conclusions: Our study shows promising results for performing neuraxial anesthesia in phantoms using the HoloLens. Although this may have wide-ranging implications for image-guided therapies, further study is required to quantify the accuracy and safety benefit of using holographic guidance.

Trial Registration: ClinicalTrials.gov NCT04028284; https://clinicaltrials.gov/ct2/show/NCT04028284

(JMIR Form Res 2022;6(6):e36931) doi:10.2196/36931

KEYWORDS

mixed reality; virtual reality; augmented reality; HoloLens; holograms; neuraxial anesthesia

Introduction

Neuraxial anesthesia has traditionally been a landmark-based technique, relying on operator feel, skill, and experience. Difficulty is highly influenced by patient body habitus, where obese patients or anatomical variations such as scoliosis or osteophytes increase difficulty and result in a higher failure rate [1,2]. Neuraxial anesthesia is not a benign procedure as multiple attempts or inaccurate needle trajectories can be anxiety provoking, cause patient discomfort, and lead to morbidity in the form of spinal/epidural hematomas, infection, dural puncture with a high risk of subsequent headaches, and nerve injury [3].

Though increasing in popularity, ultrasound guidance for neuraxial procedures is still relatively uncommon secondary to technical challenges of real-time guidance in conjunction with the difficulties of ultrasound imaging of bony structures [4]. In contrast to ultrasound-guided peripheral nerve blocks done with real-time guidance where the needle tip is visualized, the common technique for ultrasound use in neuraxial anesthesia is to provide preprocedure landmarks so the operator estimates the placement of the needle tip, depth, and trajectory before needle insertion. Anatomical landmarks are visualized using the ultrasound along multiple viewing planes and skin markings are made based on these images. The ultrasound probe is then removed from the site, placed at rest and subsequent needle insertion is done blindly based on the skin markings and the provider's recollection of approximate depth and trajectory from memory. This is not true ultrasound guidance per se, but rather ultrasound-assisted guidance. Little is known about the accuracy with which operators replicate an ideal needle path once identified via ultrasound.

The Microsoft HoloLens was introduced in 2016 and is the first self-contained, head-mounted mixed reality (MR) computing device. The headset is equipped with 4 tracking cameras and an infrared time-of-flight sensor, which allow 3D mapping of the surrounding environment, objects, and the user's hands in real time. It allows for MR, positionally stable holograms projected into a user's visual field. The user interface allows the detection of intuitive hand gestures or voice commands for application control [5]. Medical applications have included education, remote consultation, preoperative surgical planning, and surgical/procedural navigation [6]. We aimed to develop a proof-of-concept MR solution using the HoloLens to aid neuraxial blockade by allowing visualization of the ideal needle path identified on preprocedure ultrasound.

Methods

Ethical Considerations

This pilot study is part of a randomized controlled trial titled "Using Augmented Reality to 3D Map Needle Pathways in Real Time to Enhance Neuraxial Anesthesia," which has been approved by the Sunnybrook Research Institute Ethics Board (#291-2018). Study objectives and protocol were explained in detail to eligible participants (anesthesia residents and attendings), after which both verbal and written consent were obtained.

Hardware Configuration and Software Development

To enable the HoloLens to detect the ultrasound transducer position, a quick response code optical tracking marker was developed (Figure 1). First, a high-resolution 3D scan of a curvilinear ultrasound probe was carried out with an EinScan HX handheld 3D scanner (SHINING 3D Tech Co Ltd). The resulting model was modified via Autodesk Fusion 360 (Autodesk Inc., San Rafael, United States of America), a computer aided design software to create a probe mount which matched the shape of the US transducer and allowed the attachment of a 10 cm by 10 cm two-dimensional barcode. We found this to be the smallest size barcode that was reliably registered by the HoloLens's 2-megapixel camera [5]. The marker was then 3D-printed via a CR-10S 3D printer (Creality) using polylactic acid material.

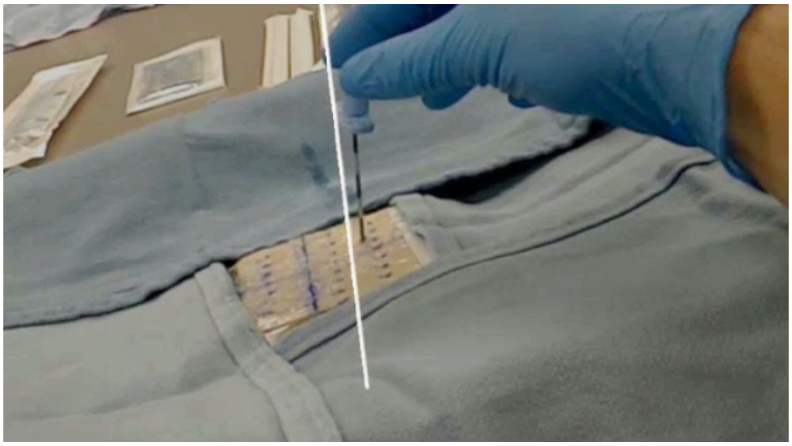
Software development was conducted in the Unity development environment (Unity Technologies). The software will be made open source following completion of future studies. Detection of the 2D barcode was accomplished by incorporating the Vuforia Augmented Reality (AR) Software Development Kit (PTC Inc). The software allows the HoloLens headset to precisely register the position of the barcode, and therefore the ultrasound transducer, via the tracking marker. Upon registering the ultrasound marker location, the operator then confirms on the ultrasound screen that the desired trajectory is displayed. The trajectory hologram is frozen by putting the ultrasound probe and marker outside the HoloLens camera visual field. We have found that the ultrasound probe marker's position is registered by the HoloLens in most configurations that are ergonomic for ultrasound use. A hologram of a line is then projected into the headset user's visual field in the location of the central axis of the ultrasound transducer (Figure 2). The position, angulation, and size of the needle path hologram remain constant as the HoloLens operator moves. Likewise, if the patient moves, the needle path hologram does not change position and hence the patient must return to their original position to maintain the accuracy of the previously identified ideal needle path. Of note, the hologram does not provide any visual projection of the optimal depth.

To establish feasibility, a lumbar spine neuraxial phantom was created for pilot testing of the guidance system [7]. Five attending anesthesiologists and two anesthesia residents were recruited. They were given a 5-minute orientation to the developed MR needle guidance system. After the orientation, we allowed an unlimited amount of time to practice using the HoloLens technology on lumbar spine neuraxial phantoms.

Multimedia Appendix 1 provides an in-depth look at how the HoloLens is used for guidance of needle angulation in phantom models. The difference between the Tuohy needle and hologram positions is due to a recording artifact and is not perceived by the user.

Figure 1. Optical tracking marker attached to a curvilinear probe to enable the HoloLens to detect transducer position.

Figure 2. This figure depicts a HoloLens user's view of aligning a Tuohy needle with a hologram representing an ideal needle path (white line) into an ultrasound phantom. The difference between the Tuohy needle and hologram positions is due to a recording artifact and is not perceived by the user.



Augmented Procedural Technique Using MR

Similar to established techniques, the patient is optimally positioned for neuraxial anesthesia. The sitting or lateral decubitus position may be used. A patient positioning device is ideal to minimize patient movement. The operator performs the procedure while wearing the HoloLens headset, which minimally interferes with procedure ergonomics and visibility. A preprocedural, nonsterile ultrasound scan of the lumbar spine is performed with the prepared tracking marker attached to the transducer (Figure 3). The posterior complex is identified and placed in the middle of the ultrasound screen. The angulation of the ultrasound transducer is then registered by the headset detecting the position of the ultrasound transducer marker, and a hologram is projected into the user's workspace, which replicates the needle path through the middle of the transducer (Figure 3 inset) in a clear path from skin to posterior complex. The central axis of the ultrasound transducer is represented by a holographic white or orange line (10 mm length, 2 mm diameter). These steps can be adapted for a paramedian approach.

The transducer is removed from the field, and its center point is marked on the skin in the usual fashion. The spatially stable hologram representing the ideal needle path remains projected into the user's visual field although the hologram does not provide any information on the desired needle depth nor does it register the needle's position in the user's visual frame. Typical sterile prepping and draping and local anesthetic injection do not disrupt the position of the needle path hologram. Operators may then use this hologram to precisely align the needle angulation with the holographic projection in 3 dimensions from the marked skin entry point (Figure 4).

Figure 3. The user performs a preprocedural ultrasound, allowing the headset to subsequently generate a spatially stable hologram. The inset shows the operator's mixed reality view displaying a holographic guidance graphic (orange line along the central axis of the ultrasound transducer) with the tracking marker attached to the ultrasound probe.

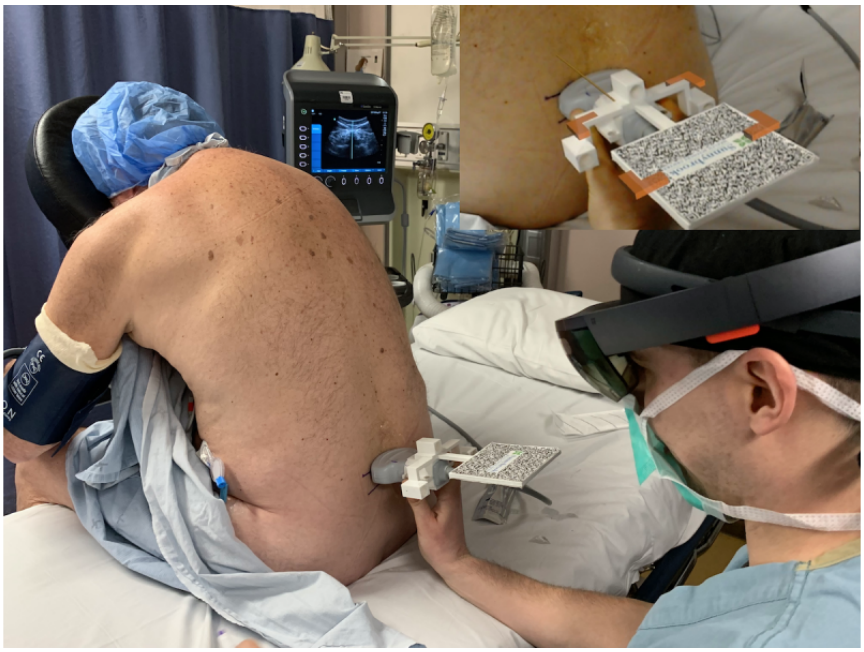
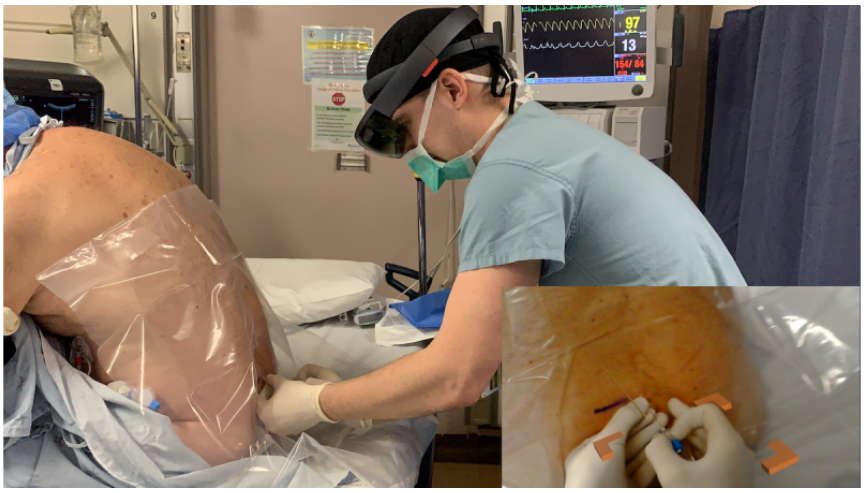


Figure 4. The user performs a neuraxial technique in standard sterile fashion. The inset shows the hologram projected into the user's visual field to aid replication of the ideal needle path as identified on ultrasound.



Results

After an informed consent process, 7 participants were recruited for the study. Five were attending anesthesiologists with a minimum of 2 years of clinical experience as consultants and two were anesthesia residents (postgraduate year 3 or above). Upon completion of the orientation and practice session, participants were asked a yes or no question: “Do you feel adequately prepared and comfortable to use this needle guidance system with a patient?” All participants indicated “yes.” All participants required 3 or fewer practice repetitions with the lumbar spine neuraxial phantoms to feel comfortable.

Discussion

Principal Findings

To our knowledge, this study is the first description of the development and feasibility testing of an MR tool for neuraxial anesthesia using a head-mounted display. We developed holographic needle guidance software using the Microsoft HoloLens and determined its use to be feasible in lumbar spine phantoms. Since not all procedures are amenable to real-time ultrasound due to challenges of simultaneously scanning and performing the technique (such as in the case of neuraxial blockade), establishment of the optimal midline and trajectory can be crucial for procedure success. Our MR system allows for a trajectory aid in such situations.

Multiple studies support the use of MR and/or AR systems for various procedures. Ameri et al [8] designed an MR ultrasound

image-guided system specifically for internal jugular vein central line insertion. Their system provided a virtual depiction of the needle and its trajectory throughout the procedure. They found that using this system led to a higher success rate of central line insertions in phantom models compared to ultrasound guidance alone for novice users (graduate students). However, when the system was used by 25 experienced physicians (attendings, residents, and fellows from anesthesiology, critical care, and emergency medicine), there was no benefit compared to ultrasound alone.

In an interventional radiology setting, Faiella et al [9] used a similar AR navigation system for computed tomography–guided percutaneous lung biopsies. The system used separate sensors to track needle position and orientation as well as patient movement. Their group found the system easy to use and diagnostically accurate with a low complication rate. This system was particularly efficacious for pulmonary nodules less than 10 mm in size, for which they noted the most drastic reduction in procedure time and a greater proportion of histological diagnoses obtained. Also in an interventional radiology setting, Marker et al [10] successfully used an AR-navigated interventional magnetic resonance imaging system for perineural injections of the thoracic, lumbar, and hypogastric sympathetic plexi, quoting a mean needle tip error of 3.9 (SD 1.7) mm and a mean procedure time of 33 (SD 12) minutes. Similar to our study, their system provided an ideal needle trajectory but did not track the needle in real time. Cumulatively, these studies demonstrate the efficacy of MR/AR systems for multiple procedure types and suggest potential for increased procedural efficacy and efficiency.

However, various limitations exist for the use of MR/AR for medical procedures. In a study of 17 participants comparing the use of AR with ultrasound alone, there was no statistically significant difference in the accuracy of identifying spinal levels prior to epidural placement between the two modalities, suggesting that AR may not enhance procedure accuracy [11]. A 2018 study by Condino et al [12] compared 20 participants' performance on connect-the-dots tasks using the Microsoft HoloLens versus the naked eye on monocular and binocular trials. Although participants rated task workload and visual comfort as being similar between modalities, user performance was statistically superior during naked eye trials. This group concluded that AR devices may not increase the precision of manual tasks.

Other challenges presented by the use of MR/AR in clinical settings, especially for new users, include visual field distortion, attentional blindness, and challenges with software user interface manipulation (especially while conducting a sterile procedure) [13]. Cost is also a potential barrier to accessing these technologies. The most recent version of the Microsoft HoloLens is currently retailing for US $3500 [5].

Although this study shows promising results, further data are needed to investigate the effectiveness of MR use for neuraxial blockade. Our group is currently conducting a randomized controlled trial to compare traditional techniques for thoracic epidural placement to an MR/HoloLens-assisted technique for elective abdominal surgery at Sunnybrook Health Sciences Centre (ClinicalTrials.gov identifier NCT04028284).

Conclusions

In this study, we report the successful development and first use of an MR needle guidance technique for neuraxial anesthesia using a head-mounted device, the Microsoft HoloLens. Although this may have wide-ranging implications for many image-guided therapies, further study is required to quantify the potential accuracy and safety benefit of holographic guidance.

Acknowledgments

We would like to acknowledge Andrew Fleet and Lilia Kaustov for formatting the manuscript for submission.

The authors purchased equipment supported by the Sunnybrook Health Sciences Centre Alternative Funding Plan Innovation Fund and an in-kind donation of equipment and time from New Age Robotics and Controls Inc for 3D scanning. Neither entity had a role in study methods development, manuscript preparation, or approval.

Authors' Contributions

JW conceived the study and developed the software/equipment related to needle guidance. FA, CM, SC, PM, and OS participated in phantom creation, HoloLens device training, and system design guidance. JT and JW performed a literature search and were the major contributors in writing this manuscript. SC and CM edited the final version of the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Holographic guidance of needle angulation in a phantom model of the lumbar spine.
[MP4 File (MP4 Video), 8191 KB - formative_v6i6e36931_app1.mp4]

References

1. de Oliveira Filho GR, Gomes HP, da Fonseca MHZ, Hoffman JC, Pederneiras SG, Garcia JHS. Predictors of successful neuraxial block: a prospective study. European Journal of Anaesthesiology 2002;19(6):447-451. [doi: 10.1097/00003643-200206000-00008]
2. Sprung J, Bourke D, Grass J, Hammel J, Mascha E, Thomas P, et al. Predicting the Difficult Neuraxial Block. Obstetric Anesthesia Digest 1999;19(3):203-217. [doi: 10.1097/00132582-199912000-00011]
3. Horlocker TT. Complications of Spinal and Epidural Anesthesia. Anesthesiology Clinics of North America 2000 Jun;18(2):461-485. [doi: 10.1016/s0889-8537(05)70172-3]
4. Perlas A, Chaparro L, Chin K. Lumbar Neuraxial Ultrasound for Spinal and Epidural Anesthesia: A Systematic Review and Meta-Analysis. Reg Anesth Pain Med 2016;41(2):251-260. [doi: 10.1097/AAP.0000000000000184] [Medline: 25493689]
5. Microsoft. URL: https://www.microsoft.com/en-us/d/hololens-2/91pnzzznzwcp [accessed 2021-05-20]
6. Hu H, Feng X, Shao Z, Xie M, Xu S, Wu X, et al. Application and Prospect of Mixed Reality Technology in Medical Field. Curr Med Sci 2019 Mar 13;39(1):1-6. [doi: 10.1007/s11596-019-1992-8] [Medline: 30868484]
7. Mashari A, Montealegre-Gallegos M, Jeganathan J, Yeh L, Qua Hiansen J, Meineri M, et al. Low-cost three-dimensional printed phantom for neuraxial anesthesia training: Development and comparison to a commercial model. PLoS One 2018;13(6):e0191664 [FREE Full text] [doi: 10.1371/journal.pone.0191664] [Medline: 29912877]
8. Ameri G, Baxter J, Bainbridge D, Peters T, Chen E. Mixed reality ultrasound guidance system: a case study in system development and a cautionary tale. Int J Comput Assist Radiol Surg 2018 May;13(4):495-505. [doi: 10.1007/s11548-017-1665-7] [Medline: 28861693]
9. Faiella E, Frauenfelder G, Santucci D, Luppi G, Schena E, Beomonte Zobel B, et al. Percutaneous low-dose CT-guided lung biopsy with an augmented reality navigation system: validation of the technique on 496 suspected lesions. Clin Imaging 2018;49:101-105. [doi: 10.1016/j.clinimag.2017.11.013] [Medline: 29207301]
10. Marker D, U Thainual P, Ungi T, Flammang AJ, Fichtinger G, Iordachita II, et al. 1.5 T augmented reality navigated interventional MRI: paravertebral sympathetic plexus injections. Diagn Interv Radiol 2017;23(3):227-232 [FREE Full text] [doi: 10.5152/dir.2017.16323] [Medline: 28420598]
11. Al-Deen Ashab H, Lessoway V, Khallaghi S, Cheng A, Rohling R, Abolmaesumi P. An Augmented Reality System for Epidural Anesthesia (AREA): Prepuncture Identification of Vertebrae. IEEE Trans Biomed Eng 2013 Sep;60(9):2636-2644. [doi: 10.1109/tbme.2013.2262279]
12. Condino S, Carbone M, Piazza R, Ferrari M, Ferrari V. Perceptual Limits of Optical See-Through Visors for Augmented Reality Guidance of Manual Tasks. IEEE Trans Biomed Eng 2020 Feb;67(2):411-419. [doi: 10.1109/TBME.2019.2914517] [Medline: 31059421]
13. Kobayashi L, Zhang X, Collins S, Karim N, Merck D. Exploratory Application of Augmented Reality/Mixed Reality Devices for Acute Care Procedure Training. West J Emerg Med 2018 Jan 18;19(1):158-164 [FREE Full text] [doi: 10.5811/westjem.2017.10.35026] [Medline: 29383074]

Abbreviations

AR: augmented reality
MR: mixed reality

Edited by A Mavragani; submitted 21.02.22; peer-reviewed by WF Tan, M Görges, J Egger; comments to author 29.04.22; revised version received 12.05.22; accepted 15.05.22; published 23.06.22.

Please cite as:
Tanwani J, Alam F, Matava C, Choi S, McHardy P, Singer O, Cheong G, Wiegelmann J
Development of a Head-Mounted Holographic Needle Guidance System for Enhanced Ultrasound-Guided Neuraxial Anesthesia: System Development and Observational Evaluation
JMIR Form Res 2022;6(6):e36931
URL: https://formative.jmir.org/2022/6/e36931
doi:10.2196/36931
PMID:35737430

Original Paper

Remotely Delivered Behavioral Weight Loss Intervention Using an Ad Libitum Plant-Based Diet: Pilot Acceptability, Feasibility, and Preliminary Results

Christina Chwyl[1], BA, MS; Nicholas Wright[2], MBChB; Gabrielle M Turner-McGrievy[3], RD, PhD; Meghan L Butryn[1], PhD; Evan M Forman[1], PhD

[1]Center for Weight, Eating, and Lifestyle Sciences, Drexel University, Philadelphia, PA, United States

[2]Royal New Zealand College of General Practitioners, Wellington, New Zealand

[3]Department of Health Promotion, Education, and Behavior, Arnold School of Public Health, Columbia, SC, United States

Corresponding Author:
Christina Chwyl, BA, MS
Center for Weight, Eating, and Lifestyle Sciences
Drexel University
3141 Chestnut Street
Philadelphia, PA, 19104
United States
Phone: 1 267 606 0414
Email: cmc646@drexel.edu

Abstract

Background: Many traditional lifestyle interventions use calorie prescriptions, but most individuals have difficulty sustaining calorie tracking and thus weight loss. In contrast, whole food plant-based diets (WFPBDs) have previously shown significant weight loss without this issue. However, most WFPBD interventions are face-to-face and time-intensive, and do not leverage gold standard behavioral strategies for health behavior change.

Objective: This open pilot trial was the first to evaluate the feasibility of a fully featured, remotely delivered behavioral weight loss intervention using an ad libitum WFPBD.

Methods: Over 12 weeks, participants (N=15) with overweight or obesity received a newly designed program that integrated behavioral weight loss and a WFPBD prescription via weekly web-based modules and brief phone coaching calls. Assessments were performed at baseline, midtreatment (6 weeks), and after treatment (12 weeks).

Results: The intervention was rated as highly acceptable (mean 4.40 out of 5, SE 0.18), and attrition was low (6.7%). In all, intention-to-treat analyses revealed that 69% (10.4/15) of the participants lost 5% of their weight (mean –5.89, SE 0.68 kg). Predefined benchmarks for quality of life were met.

Conclusions: A pilot digital behavioral weight loss intervention with a non–energy-restricted WFPBD was feasible, and the mean acceptability was high. Minimal contact time (80-150 minutes of study interventionist time per participant over 12 weeks) led to clinically relevant weight loss and dietary adherence for most participants (10.4/15, 69% and 11.8/15, 79%, respectively), and quality of life improvements (reliable change indices >1.53). We hope that this work will serve as a springboard for future larger scale randomized controlled studies evaluating the efficacy of such programs for weight loss, dietary change, and quality of life.

Trial Registration: ClinicalTrials.gov NCT04892030; https://clinicaltrials.gov/ct2/show/NCT04892030

(JMIR Form Res 2022;6(6):e37414) doi:10.2196/37414

KEYWORDS

vegetarian diet; vegan diet; overweight; eHealth; behavioral intervention

Introduction

Background

Excess weight is a leading cause of death in industrialized countries and contributes to a wide range of health issues, including cardiovascular disease, type 2 diabetes, risk of certain cancers, and early mortality [1-4]. Standard approaches to weight loss rely on calorie prescriptions to achieve negative energy balance [5]. Calorie prescriptions, in conjunction with the provision of psychological and behavioral strategies (eg, goal setting, stimulus control, self-monitoring, problem solving, and cognitive restructuring) to facilitate lifestyle modification, are considered the current gold standard behavioral weight loss treatments (S-BTs) [6]. S-BTs produce, on average, 5% to 8% of body weight loss following intensive, year-long intervention [5]. However, one-third of individuals do not lose clinically significant levels of weight, and one-third of initial weight lost is regained in the year after treatment, with continued weight regain thereafter [7].

Overall, 2 core limitations (reliance on calorie tracking and lack of appetite control optimization) may contribute to the suboptimal outcomes of S-BT. S-BTs rely on meticulous dietary self-monitoring (ie, tracking everything consumed), which is considered the cornerstone of treatment success [5]. Indeed, thorough dietary self-monitoring is one of the strongest identified predictors of weight loss and maintenance outcomes following S-BT [8-10]. However, many participants find calorie tracking to be unappealing and arduous [11,12] and are unable to track their calorie intake consistently, accurately, and in the long term [11,13-15]. Another limitation of S-BT is that it is not optimized to address the increase in appetite that individuals face when losing weight [16]. That is, when individuals lose weight, a host of biological adaptations occur, including unfavorable changes to appetite that serve to guard against fat loss [16,17]. However, although appetite is recognized to play a critical role in governing energy intake and driving suboptimal weight loss and maintenance outcomes [18], S-BTs are not optimized to address the increase in appetite that individuals face during weight loss.

In contrast, a whole food plant-based diet (WFPBD) can produce clinically significant weight loss, health, and quality of life improvements in the absence of calorie tracking [19-22]. WFPBDs include vegetables, fruits, legumes, starches, and whole grains in minimally processed forms. This maximizes nutrient-dense, low–energy density foods, while limiting energy-dense foods (eg, processed foods and animal products). Decreasing dietary energy density with ad libitum consumption can result in weight loss while sating appetite [23-25], potentially by reducing passive intake of high-calorie or high-fat foods [26] by increasing the volume of food that cues fullness [27] or owing to the satiating effects of a high-fiber diet low in energy density [28]. Indeed, WFPBDs have been found to significantly decrease ad libitum consumption compared with other diets. For example, individuals assigned to eat WFPBD meals ad libitum ate, on average, 689 calories per day fewer than those assigned to eat animal-based, low-carbohydrate meals ad libitum [24]. Plant-based diets also merit consideration for health and environmental benefits. Regarding physical health, WFPBDs are robustly associated with reduced risk of chronic reduction, including reduced rates of cardiovascular risk [21,29,30], hypertension [21], type 2 diabetes [31], and certain cancers [32]. Regarding the environment, plant-based foods are superior to animal products in terms of land use, freshwater use, land acidification, and greenhouse gas emissions [33].

Despite the potential of WFPBDs for health and weight loss, existing interventions using WFPBDs have 2 key limitations. First, the vast majority of WFPBD interventions are conducted in-person, and delivery requires significant time of participants and highly trained professionals [7,20] (for an exception, see the study by Turner-McGrievy et al [34]), of which there is a shortage [35,36]. In contrast, remotely delivered weight loss interventions, especially those with minimal human support, have a higher potential for dissemination and can produce clinically significant weight loss when engagement (eg, interactivity) and health behavior change (eg, accountability and self-monitoring) features are included [15,37-39]. In addition, although gold standard approaches to weight loss incorporate psychological and behavioral strategies to facilitate lifestyle modification (eg, stimulus control and regular self-weighing), most WFPBD interventions have provided nutrition information alone [20,40]. This likely limits their efficacy, as behavioral and psychological skills are known to improve the efficacy of weight loss interventions [41], and simply knowing what to do to lose weight does not necessarily translate into behavior change, as demonstrated by the superior efficacy of S-BT to psychoeducation alone [42].

Objectives

This open trial pilot study addressed these limitations by evaluating the feasibility of a remotely delivered digital behavioral weight loss intervention using an ad libitum WFPBD. The intervention was designed with accessibility in mind and thus was remotely delivered and required minimal time from the participants and clinicians. The intervention consisted of two core components: (1) weekly web-based modules delivering WFPBD nutrition counseling and gold standard behavioral weight loss strategies and (2) one-on-one, 10- to 15-minute phone *coaching* calls, with a study interventionist for most weeks. The primary aim was to evaluate recruitment feasibility and acceptability. The secondary aim was to evaluate the preliminary effect of the intervention on weight loss and dietary adherence. As an exploratory aim, we evaluated the preliminary impact of the intervention on the quality of life.

Methods

Participants

Adult men and women with overweight or obesity (BMI ≥25 kg/m^2), aged 18 to 75 years, and residing in the United States were eligible as part of larger recruitment efforts for the ongoing weight loss trials at the Drexel University Center for Weight, Eating and Lifestyle Science (WELL Center). Recruitment for other weight loss trials in the center occurred through advertisements at radio stations and in newspapers and social media posts. Notably, no advertising specific to this study was

conducted. Individuals interested in participating in a weight loss study were contacted via phone, and preliminary eligibility for weight loss trials in the WELL Center was assessed. Depending on the enrollment status of the other ongoing trials, individuals were provided preliminary information about this study and could elect to be screened for it. In addition, individuals ineligible for other trials in the WELL Center could elect to be screened for this study. Trained research staff then conducted a preliminary screening for this study. Approximately 0 to 2.5 months later, the first author (CC) contacted interested participants to discuss full eligibility. Full eligibility was assessed by phone, and baseline assessments were scheduled for those who were interested and eligible. Figure S1 in Multimedia Appendix 1 details participant flow and reasons for exclusion. Of the individuals who underwent preliminary screening for this study (N=86), the reasons for exclusion included lack of interest in following the study diet (25/86, 29%), >5% weight loss in the past 3 months (10/86, 12%), medical condition influencing weight or appetite (8/86, 9%), ongoing participation in another weight loss trial underway at the center (6/86, 7%), eating disorder pathology (3/86, 3%), poor English comprehension (3/86, 3%), already following the study diet (2/86, 2%), inability to attend appointments (2/86, 2%), BMI <25 (2/86, 2%), inability to follow the study diet (1/86, 1%), and consumption of medication known to cause weight gain (1/86, 1%).

The participants were recruited between November 2020 and February 2021. Given the novelty of the intervention, before the pilot trial, 7 participants were enrolled in a pretest of the intervention between December 2020 and February 2021. Following the pretest, refinements to the intervention were made, including streamlining content, addressing common confusion about the WFPBD earlier on, and adding more structured questions and goal setting to the phone coaching calls. Table S1 in Multimedia Appendix 1 presents a list of intervention refinements and the rationale for each. The refined intervention was then delivered to a set of 15 participants in the pilot trial, starting between January and February 2021 and ending between March and April 2021. Because this was a pilot feasibility study, a power analysis was not performed [43]. We aimed for a sample of 14 participants, because we deemed this sample size sufficient to evaluate feasibility and acceptability. Notably, this sample size is consistent with prior internet-delivered weight loss pilot interventions [44]. Exclusion criteria were the use of medications for weight loss, ≥5% weight loss in the past 3 months, current or planned pregnancy within the study period, bariatric surgery history, currently following a WFPBD, diagnosis of a serious medical or psychiatric condition influencing weight or appetite, high substance use, and eating pathology (ie, binge eating disorder diagnosis or subthreshold loss-of-control eating or compensatory behaviors).

Study Design

The Template for Intervention Description and Replication checklist and guide was used to inform the description of the intervention [45]. Reporting followed the guidelines of the CONSORT (Consolidated Standards of Reporting Trials) extension to pilot and feasibility trials (excluding items specific to randomization) [46]. The intervention comprised two components: (1) weekly web-based modules delivering behavioral weight loss strategies and WFPBD nutrition psychoeducation and (2) brief 10- to 15-minute *coaching* calls with a study interventionist.

Dietary Prescription

Participants were prescribed a WFPBD that promoted the intake of fruits, vegetables, starches, legumes, and whole grains [47]. Participants were encouraged to avoid processed foods, refined oils, and animal products and to minimize the consumption of high-fat plant-based foods. An adapted *traffic-light* diet chart (Figure S2 in Multimedia Appendix 1 [47]) outlined foods to eat, limit, and avoid. The participants were advised to eat until satiation (without restricting energy intake). Consistent with past WFPBD interventions and to ensure nutritional adequacy, participants were instructed to consume 50 µg of vitamin B12 supplements daily [47,48]. The participants were asked to purchase vitamin B12 supplements. If doing so imposed a financial burden, participants were mailed vitamin B12 supplements.

Behavioral Intervention

The behavioral or psychological strategies used in the intervention were developed based on existing gold S-BT [49], evidence-based behavioral change techniques (BCTs) [50], and social cognitive theory [51]. Content specific to WFPBD was primarily adapted from the BROAD study [47]. The intervention was also informed by feedback from 7 participants who experienced the preliminary *beta* version of the program (in particular, participant feedback resulted in the addition of mobile-friendly and printable PDFs of key materials, links to transcripts of videos, individualized weight goals, streamlined modules, more detailed information on the traffic-light diet chart provided in weeks 1 and 2, increased representation of body sizes in intervention content, a Google Classrooms tutorial, and guided inquiry related to nonadherence to the WFPBD). To inform intervention replication, Table S2 in Multimedia Appendix 1 presents a list of the BCTs used in this study according to the BCT taxonomy by Michie et al [50], which includes 93 individual behavior techniques grouped into 16 superordinate clusters. The intervention included a variety of BCTs (14 BCT clusters and 31 specific techniques).

Modules

The weekly web-based modules delivered in the intervention were created using a Google Slide add-on (Pear Deck) that enabled the integration of interactive questions were hosted on a popular e-learning platform (Google Classroom). Content was presented through a combination of audio, text, images, and interactive elements (eg, written reflection prompts, draggable questions, and *Knowledge Check* quizzes). The materials were adapted from existing, successful, and in-person behavioral weight loss treatments. Specifically, WFPBD nutrition counseling content was based primarily on materials delivered in the BROAD study [47] and featured materials on energy density [25], food cravings [52], and nutrition psychoeducation [53,54]. Psychological and behavioral skills were similar to those used in the Diabetes Prevention Program protocols [55] and past successful behavioral weight loss treatments [56] and

included core behavioral strategies for weight loss and maintenance, such as stimulus control [49], problem solving [49], habit formation [57], and relapse prevention [58]. Content was divided into three parts: (1) transition (weeks 1 and 2, in which participants were asked to begin transitioning to a WFPBD), (2) change (weeks 3-6, in which participants were asked to fully adopt a WFPBD), and (3) sustain (weeks 7-12, in which participants received behavioral and psychological skills aimed at facilitating long-term weight maintenance). The modules were designed to take between 10 and 30 minutes to complete. Table S3 in Multimedia Appendix 1 presents an outline of the weekly intervention content. Following each module, to solidify learning, participants completed weekly worksheets (*Put into Practice Assignments*).

Each week, a new module was made available. The participants accessed each module individually and completed it at their own pace. Participants could complete modules at any time within the assigned week but were encouraged to set aside a consistent time to view the intervention content each week. Each week, participants had access to additional recipes, optional further reading, and examples of people who had successfully adopted a WFPBD lifestyle. We included cooking and educational resources from a diverse set of people in terms of age, race, gender, and cultural background. If participants did not complete the module, up to 2 follow-up emails were sent or reminder calls, made.

Phone Coaching

To facilitate dietary adherence and provide accountability and individualized feedback, key components of successful weight loss interventions [50], 9 to 11 one-on-one, 10- to 15-minute phone check-in calls were scheduled. The participants were permitted to miss up to 2 coaching calls in the event of holidays or sickness, and the final coaching call was optional. Phone coaching was conducted by the primary investigator, a clinical psychology doctoral student with training and prior experience in delivering behavioral weight loss. Before the study, the primary investigator received day-long training in behavioral weight loss from experienced, advanced practitioners and received weekly individual supervision on effective delivery of behavioral weight loss. Calls consisted of positive reinforcement, problem-solving support, and motivational support, in line with a motivational interviewing approach [59]. Figure S3 in Multimedia Appendix 1 details the full phone coaching protocol.

Weekly Weighing

Given the research on the advantages of frequent self-weighing [60], each week, participants entered their weight into a web-based spreadsheet that auto-populated a graph to visually depict progress. Automated emails prompted participants to enter their weight.

Measures

Assessments were performed at baseline, midtreatment (6 weeks), and after treatment (12 weeks) and were conducted remotely via Zoom (Zoom Video Communications, Inc). Informed consent was obtained at the baseline assessment, and participants were provided a tutorial on how to complete the modules and use Google Classrooms. A compensation of US $10 for the midtreatment (6 weeks) and US $20 for the posttreatment (12 weeks) assessment was provided.

Plant-Based Diet Familiarity

To contextualize our sample, we assessed plant-based diet history at baseline by asking participants to report whether they had eaten a plant-based diet in the past, for at least three months. Participants' responses were coded into the following categories: previously vegan, vegetarian, pescatarian, flexitarian, or no prior plant-based diet familiarity (omnivorous).

Acceptability

Acceptability was assessed after treatment with a questionnaire adapted from prior work [44] that asked participants to report how satisfied they were with the program, the degree to which they found the program helpful, and the likelihood that they would recommend it to family or friends using a Likert scale ranging from 1 (not at all) to 5 (very much). After treatment, participants reported how helpful they found each intervention component (phone check-ins, modules, Put into Practice Assignments, and weight charts). Items were averaged to calculate a composite acceptability score. The intervention was considered feasible if the mean acceptability ratings >4 out of 5 for at least 80% of the participants, a rate consistent with prior digital interventions [44]. To inform iterative development, following each module, participants rated the degree to which the module was helpful and engaging using the same scale. Modules with an average rating >4 out of 5 were considered acceptable.

Feasibility

To evaluate feasibility, we examined the percentage of participants who were successfully retained in the study (defined as completing at least 10 of the 12 weekly modules and posttreatment assessment). The intervention was considered feasible if at least 80% of the participants met our retention criteria (completing at least 10 of the 12 modules as well as the posttreatment assessment), a retention rate consistent with previous brief digital interventions [39]. In addition, we examined the number of modules completed by the participants.

Anthropometric Data

Participants self-weighed weekly and at the baseline, midtreatment (6 weeks), and posttreatment (12 weeks) assessments, following recommended guidelines, that is, instructed to weigh the first thing in the morning at least three times (a fourth if weights are discrepant by >0.2 lbs), in no or light clothes, and ensuring the scale is on a hard, flat surface. Weight was self-reported in pounds and was later converted to kilograms by the researchers. Self-reported weight following such guidelines generally have high accuracy [61,62]. Consistent with the weight losses observed in past interventions of a 12-week duration that were successfully delivered remotely with minimal human support [39,63,64], we considered weight loss to be suggestive of a preliminary impact if approximately 50% of our sample lost clinically meaningful levels of initial body weight (5% of initial body weight) [65].

Dietary Adherence

At baseline, midtreatment (6 weeks), and after treatment (12 weeks), participants completed an adapted 18-item food frequency questionnaire aligned with the prescribed WFPBD traffic-light diet (Figure S2 in Multimedia Appendix 1 [47]) using a Likert scale ranging from 1 (rarely or never) to 6 (3 or more times per day). We examined the mean differences in consumption of green-, yellow-, and red-zone foods and overall dietary improvements (ie, weighted scores for the yellow- and red-zone foods subtracted from the green-zone foods). To weight scores, we multiplied responses in the ultragreen- and ultrared-zone categories by 3 and light green– and light red–zone food categories by 2. For interpretability, we transformed the scores to a 0 to 100 scale, with higher scores representing greater adherence. We considered dietary change to be meaningful and indicative of a preliminary impact if at least 80% of participants improved their dietary adherence score by at least 20 points. Adapted food frequency questionnaires have been found to be a valid method for capturing dietary intake across a wide range of omnivorous and plant-based diets [66].

Quality of Life

Participants completed the 36-item Short-Form General Health Survey (SF-36) [67] at baseline, midtreatment (6 weeks), and after treatment (12 weeks). The SF-36 assesses 8 domains: physical functioning, limitations owing to physical health, pain, general health, energy or fatigue, social functioning, emotional well-being, and mental health. This measure has been shown to have good validity [68]. Scores were converted to a 0 to 100 scale, with higher scores indicating better functioning. This scale produces summaries of both physical health and mental health components. Owing to an error in survey creation, item 22 was not presented at baseline, although it was presented at midtreatment and after treatment. The omission of item 22 did not appear to change the results or adversely affect internal consistency. Internal consistency in our sample was excellent for physical health (α=.92) and good (α=.85) for the mental health component summary. Because the level of change in SF-36 scores that reflect a clinically significant improvement among weight loss samples is not yet well established, we calculated a reliable change index at the 0.20 level [69]. If the product exceeds a *z*-score of 1.28, reflecting 80% confidence, we consider our results to be suggestive of a preliminary impact.

Statistical Analyses

Descriptive statistics were calculated for each outcome and process measure to evaluate whether benchmarks (specified a priori) thought to represent feasibility across several outcome measures were achieved (Table 1). Wilcoxon signed-rank tests were used to examine the preliminary impact of the intervention on weight loss, dietary change, and quality of life from before to after treatment. All variables are reported as the mean and SE of the mean or as frequencies. Analyses were conducted per-protocol (PP) and using the intention-to-treat (ITT) approach [55]. We imputed missing data (n=1 at midtreatment and after treatment) using multiple imputation procedures, a recommended approach for dealing with missing weight loss data [70]. We used SPSS (version 26) to conduct multiple imputations, specifying 5 generated data sets.

Table 1. Summary of intervention outcomes.

Outcome	Benchmark	Benchmark attainment
Feasibility	≥80% retained	93.3[a]
Program acceptability	≥80% acceptability ≥4 out of 5	77.3
Weight loss	≥50% achieving 5%	69.3[a]
Dietary adherence	≥80% changed diet by ≥20 scale points	78.7
Quality of life, RCI[b]		
Physical health component summary	≥1.28	3.94[a]
Mental health component summary	≥1.28	1.53[a]

[a]Outcomes that met or exceeded our benchmark. The results were reported from the intention-to-treat analyses.

[b]RCI: reliable change index.

Ethics Approval

All aspects of the study design were specified a priori and the Drexel University Institutional Review Board approved this study (protocol number 2008008061). The study was registered at ClinicalTrials.gov on May 19, 2021 (identifier NCT04892030). All the participants provided written informed consent.

Results

Baseline Characteristics

A total of 15 individuals participated in the study. The sample was primarily White (11/15, 73%; 3/15, 20% Black or African American; 1/15, 7% South Asian), middle-aged (mean 52.67, SE 2.12 years), employed full- (11/15, 73%) or part-time (1/15, 7%), and married (10/15, 67%). All participants were female and identified as women. BMI ranged from 26.66 to 80.49 kg/m^2 (median 35.25, SE 4.58 kg/m^2). All participants endorsed at least some higher education: with 47% (7/15) reporting a

postgraduate degree, 27% (4/15) reporting a college degree, 13% (2/15) reporting an associate degree, and 13% (2/15) reporting some college. Participants had variable prior experiences with plant-based eating: with 13% (2/15) having been flexitarian, 13% (2/15) having been pescatarian, 20% (3/15) having been vegetarian, and none having been vegan. There was no discernible pattern for income brackets, and 20% (3/15) of the participants chose not to answer.

Summary of Results

Table 1 summarizes the benchmark attainment for each outcome measure. Table 2 presents the descriptive statistics of each outcome measure at midtreatment and after treatment, separately. Figure 1 presents the outcomes at midtreatment and after treatment. Variables are reported as mean and SE or as frequency and percentage.

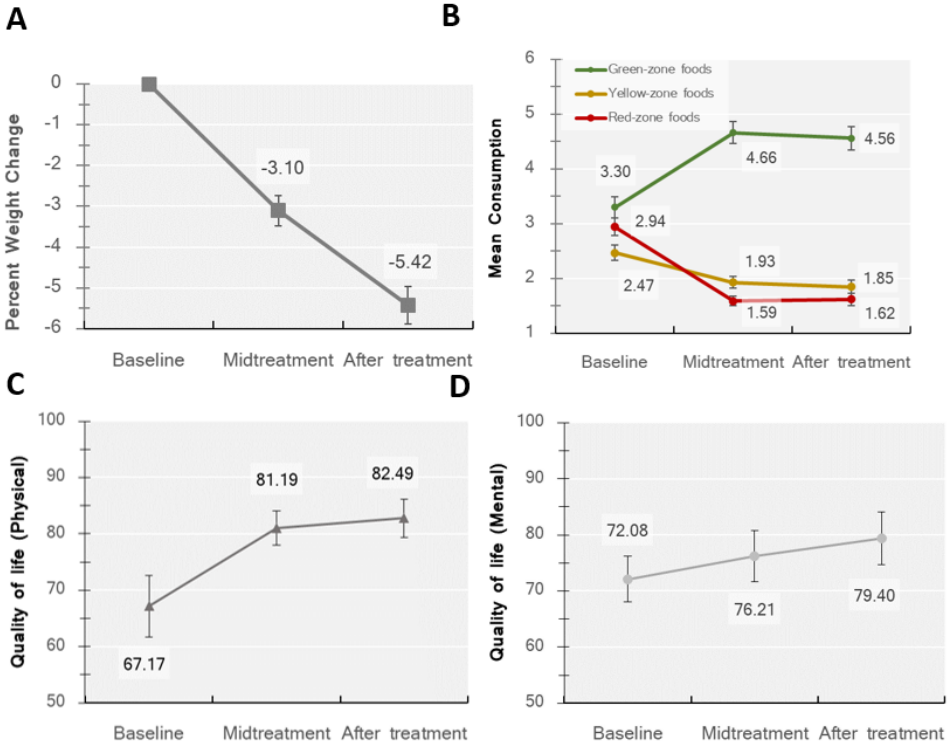
Table 2. Differences in outcome measures at midtreatment and after treatment.

	Baseline (0 weeks), mean (SE)	Midtreatment (6 weeks), mean (SE)	After treatment (12 weeks), mean (SE)	Change, mean (SE)
Anthropometric measures				
Weight (kg)	113.95 (11.87)	104.67 (10.09)	102.03 (9.97)	−5.89 (0.68)
Self-report surveys				
WFPBD FFQ[a] (0-100)[b]	56.24 (1.95)	82.35 (2.01)	81.35 (2.21)	25.17 (1.70)
Green zone (raw mean)	3.30 (0.19)	4.66 (0.20)	4.56 (0.21)	1.14 (0.14)
Yellow zone (raw mean)	2.47 (0.14)	1.93 (0.11)	1.85 (0.13)	−0.61 (0.11)
Red zone (raw mean)	2.94 (0.16)	1.59 (0.09)	1.62 (0.11)	−1.38 (0.12)
Quality of life (physical)	67.17 (5.45)	81.02 (3.08)	82.76 (3.38)	13.54 (3.96)
Quality of life (mental)	72.08 (4.07)	76.21 (4.52)	79.40 (4.73)	7.43 (4.78)

[a]WFPBD FFQ: whole food plant-based diet food frequency questionnaire.

[b]The WFPBD FFQ transformed to a 0 to 100 scale. A score of 0 represented complete dietary nonadherence (ie, frequent intake of red and yellow zone foods and no or limited intake of green-zone foods), and a score of 100 represented complete adherence.

Figure 1. Outcomes at midtreatment and after treatment. Mean values for participants (N=15) at baseline, midtreatment (6 weeks), and after treatment (12 weeks) for (A) percent weight change, (B) diet change on the adapted food frequency questionnaire, (C) quality of life (physical component) as assessed by the 36-item Short-Form General Health Survey (SF-36), and (D) quality of life (mental component), as assessed by the SF-36. Error bars represent SE of the mean. Percent weight change and percent waist change, rather than percent weight loss and waist circumference loss, are depicted by the convention.



Feasibility

The feasibility benchmark of retaining at least 80% of the participants was met (Table 1). A participant dropped out of treatment in week 2. Of the remaining participants, only 1.2% (2/168) modules were collectively missed. Participants received 8 to 10 coaching calls (80-150 minutes of coaching over the 12-week intervention period). Of the treatment completers, 14% (2/14) received 8 calls, 36% (5/14) received 9 calls, and 50% (7/14) received 10 calls. The reasons for missed calls included holidays and sickness.

Acceptability

Acceptability was, on average, high (ITT: mean 4.40, SE 0.18; PP: mean 4.43, SE 0.18), although the percentage of participants attaining the program acceptability benchmark (ITT: mean 77.3%, SE 11.9%; PP: mean 73.3%, SE 26.7%) fell slightly below our goal of 80% attaining this benchmark (Table 1). Participants rated the weekly weigh-ins as most helpful (ITT: mean 4.53, SE 0.19; PP: mean 4.57, SE 0.17), followed by the web-based modules (ITT: mean 4.39, SE 0.21; PP: mean 4.43, SE 0.2), phone check-ins (ITT: mean 4.2, SE 0.29; PP: mean 4.21, SE 0.28) and Put into Practice Assignments (ITT: mean 3.96, SE 0.34; PP: mean 4, SE 0.31). All modules were rated as acceptable (≥4 out of 5) except for week 7, which presented content on common nutrition myths and advanced cooking techniques (ITT: mean 3.87, SE 0.27; PP: mean 3.89, SE 0.27) and week 8, which presented content on social support (ITT: mean 3.97, SE 0.23; PP: mean 3.96, SE 0.23). The highest rated module was week 11, which presented content on relapse prevention (ITT: mean 4.45, SE 0.19; PP: mean 4.42, SE 0.21), followed by week 9, which presented content on stress and emotional eating (ITT: mean 4.42, SE 0.21; PP: mean 4.46, SE 0.2).

Weight Loss

The weight loss benchmark was attained (Table 1). Weight loss (ITT: mean 5.89, SE 0.68 kg; PP: mean 5.86, SE 0.73 kg) was observed before to after intervention (z=−3.41; P<.001; Cohen d 0.74).

Dietary Change

The percentage of participants meeting the dietary adherence benchmark (ITT: 11.8/15, 79%; PP: 11/14, 79%) fell slightly short of our goal of 80% of participants meeting this benchmark (Table 1). Large dietary changes (ITT: mean 25.17, SE 1.70; PP: mean 25.2, SE 1.80) were observed from before to after treatment (z=−3.41; P<.001; Cohen d 1.55).

Quality of Life

The quality of life benchmark was met (Table 1). Large changes in the physical health component summary of the SF-36 (ITT: mean 13.54, SE 3.96; PP: mean 13.48, SE 4.24) were observed before to after treatment (z=−3.24; P<.001; Cohen d 0.97). Small to medium changes (ITT: mean 7.43, SE 4.78; PP: mean 7.59, SE 5.13) were observed in the mental health component summary (z=−1.24; P=.22; Cohen d 0.45).

Discussion

Principal Findings and Comparison With Prior Work

This open pilot trial was the first to evaluate a remotely delivered ad libitum WFPBD behavioral weight loss intervention for adults with overweight or obesity, with minimal coaching support contact (80-150 minutes per patient over the course of 3 months). Our study was unique in that it integrated behavioral weight loss with an ad libitum WFPBD prescription that was remotely delivered. The results support the feasibility of the intervention and the preliminary impact of the intervention on weight loss and quality of life. The overall acceptability ratings were high, although they did not reach our acceptability benchmark (ITT: 77.3% vs the 80% prespecified). Similarly, dietary change over the course of the intervention was large (Cohen d 1.55), although dietary adherence did not reach the dietary adherence benchmark (ITT: 78.7% vs 80% prespecified). Thus, this study extends research on the feasibility and, potentially, the acceptability of a remotely delivered lower intensity format [20]. Our results support future studies examining the impact of interventions on weight loss, dietary adherence, and quality of life.

The results are promising given the need for more accessible and lower intensity weight loss treatments [71], especially those that do not require substantial participant time (eg, calorie tracking) or staff time (approximately 10-15 minutes weekly). In contrast, current gold standard behavioral treatments are expensive to deliver and require a workforce of expert clinicians, of which there is a shortage [72]. Therefore, if supported by further research, policy makers and practitioners could offer such a program to individuals seeking weight loss.

In contrast to traditional weight loss approaches that rely on calorie tracking to achieve a negative energy balance, the outcomes of our study were achieved with an ad libitum diet. Thus, our work extends the literature on weight loss approaches that rely on natural satiation mechanisms to achieve negative energy balance [25,73]. Although our results bear replication, the percentage of participants meeting the 5% weight loss threshold is similar to the level of weight loss achieved in other web- or app-based behavioral weight loss programs requiring food tracking and supplemented by remote human support [39,74]. In addition, clinically significant changes in diet and quality of life were observed in most participants. Thus, in contrast to prevailing weight loss approaches that produce weight loss to the extent that individuals track calories and adhere to specific calorie goals [5], WFPBDs do not require burdensome calorie tracking. Instead, they allow individuals to achieve negative energy balance through natural satiation mechanisms and may thus represent a promising alternative weight loss approach. Further research on behavioral weight loss treatments using ad libitum diets is critical, given the demand for non–calorie-tracking weight loss approaches [12,75] and given that calorie tracking noncompliance is prevalent and largely tracks weight regain [7,9] among individuals in weight loss programs that prescribe calorie goals.

Finally, this study was unique in its integration of WFPBD nutrition psychoeducation with behavioral and psychological

principles for health behavior change [49,50,55]. Such strategies (eg, regular self-weighing and relapse prevention), as well as the inclusion of features shown to produce engagement and health behavior change in prior remotely delivered interventions (eg, interactivity, personalized elements, accountability, and self-monitoring), may facilitate efficacy [15,37,39]; however, this requires testing with a larger sample.

Limitations

This study had several limitations. First, the observed enrollment rate of 27% is somewhat smaller than that observed in other trials at our center (33%-38%) [49,76]. Several factors may have resulted in a slightly lower enrollment yield rate, including the fact that participants were recruited from a generic pool of participants (vs the usual practice of advertising a specific study), interested individuals were contacted after up to a 2.5-month delay, and screenings for several weight loss trials occurred concurrently, leading some individuals to, instead, enroll in other trials. We anticipate that recruiting for a fully powered trial will be feasible, especially with study-specific advertisements.

Dietary acceptability also affected the enrollment rate; 29% (25/86) of individuals screened for this study were excluded owing to a lack of interest in following the WFPBD, suggesting that a WFPBD is not universally acceptable. Notably, conventional diets used in S-BT (which require meticulous calorie tracking) are likewise not universally acceptable [12] and may prevent individuals from enrolling in S-BT. In addition, many individuals enrolled in S-BT are unable to sustain calorie-tracking requirements, suggesting that conventional diets are not feasible over the long term for many [15]. Thus, although not universally acceptable, ad libitum WFPBDs may represent a viable alternative, especially for those who find calorie-tracking approaches unappealing or unsustainable. A promising future research direction is to evaluate recruitment feasibility for behavioral weight loss programs using standard versus ad libitum WFPBDs.

As this was an open pilot trial, results need to be replicated in larger samples, with a control condition and long-term follow-up period. Given that ad libitum WFPBD interventions rely on natural satiation mechanisms rather than calorie tracking to achieve negative energy balance, it is possible that a behavioral weight loss program using an ad libitum WFPBD may produce more sustained weight loss outcomes than behavioral weight loss interventions using a traditional calorie-prescribed diet [77]; however, this question requires empirical testing. Given the notorious challenge of obtaining accurate self-reports of dietary intake [11], future research should include more rigorous measures of dietary intake, such as 24-hour recalls administered by a registered dietician. The dietary adherence measure in this study was likely limited by retrospective bias (eg, reports over the past month) and may have been affected by participant characteristics (eg, those with higher conscientiousness or nutrition knowledge may have had higher accuracy).

Notably, our sample consisted of middle-aged women with higher than average educational attainment [78]. The degree to which our findings can be generalized to samples beyond those represented in this study is unknown. Our sample may have also had higher than average levels of motivation to lose weight or change their diet, including greater willingness to make substantive dietary changes. Future research would benefit from examining not only the efficacy of behavioral weight loss treatments using WFPBDs but also its effectiveness in community samples [47]. It is also important to note that this study was conducted during the COVID-19 pandemic, which may have influenced dietary behavior.

Finally, to assist in intervention streamlining and cost-effectiveness, future research would benefit from disentangling active treatment components from inert ones, examining the optimal dose of costly intervention components (ie, phone coaching) and exploring stepped care approaches (eg, providing phone coaching only to those who do not respond to modules alone or automated messages). Indeed, in this intervention, acceptability ratings for phone coaching were variable (mean 4.20, SE 0.29), with some participants reporting that phone coaching was an essential treatment component for them, while others reporting a desire for less frequent meetings.

Conclusions

In sum, the results supported the feasibility of a 12-week remotely delivered intervention, prescribing a nonenergy-restricted WFPBD with minimal human contact (10-15 minutes most weeks). Feasibility was achieved, and the results support further research to evaluate efficacy for weight loss, dietary adherence, and quality of life. The program appears promising, given the need for more accessible alternatives to in-person and calorie tracking–based weight loss approaches. If supported by further research, such an intervention could be considered a frontline treatment, owing to the cost-effectiveness of digital treatments [74]; the limited time required of participants and staff; individuals' preferences for lower intensity treatments [79]; the need for remotely delivered treatments, especially in light of the COVID-19 pandemic; and the wide-ranging health benefits of this dietary approach. This intervention might be of particular benefit to individuals who do not have access to in-person treatment, who do not benefit from S-BT, or who find calorie-tracking approaches to weight loss difficult to sustain. To build upon the current research, we are currently conducting a randomized controlled trial evaluating the efficacy of a remotely delivered behavioral weight loss intervention using an ad libitum WFPBD, in comparison with the current gold standard behavioral weight loss approach (S-BT).

Acknowledgments

The authors thank Madelyn France for her assistance in conducting the assessments and preparing the intervention materials.

Conflicts of Interest

EF serves on the Scientific Advisory Board of Tivity Health. EF and MB receive royalties from Oxford University Press for a published acceptance-based treatment manual. GT-M received speaker honorarium from 2 nonprofit conferences, the Plant-Based Nutrition Healthcare Conference and the Plant-Based Prevention of Disease conferences.

Multimedia Appendix 1
Untitled.
[DOCX File , 333 KB - formative_v6i6e37414_app1.docx]

References

1. Al-Goblan AS, Al-Alfi MA, Khan MZ. Mechanism linking diabetes mellitus and obesity. Diabetes Metab Syndr Obes 2014 Dec 4;7:587-591 [FREE Full text] [doi: 10.2147/DMSO.S67400] [Medline: 25506234]
2. Global BMI Mortality Collaboration, Di Angelantonio E, Bhupathiraju S, Wormser D, Gao P, Kaptoge S, et al. Body-mass index and all-cause mortality: individual-participant-data meta-analysis of 239 prospective studies in four continents. Lancet 2016 Aug 20;388(10046):776-786 [FREE Full text] [doi: 10.1016/S0140-6736(16)30175-1] [Medline: 27423262]
3. Lavie CJ, Milani RV, Ventura HO. Obesity and cardiovascular disease: risk factor, paradox, and impact of weight loss. J Am Coll Cardiol 2009 May 26;53(21):1925-1932 [FREE Full text] [doi: 10.1016/j.jacc.2008.12.068] [Medline: 19460605]
4. Vucenik I, Stains JP. Obesity and cancer risk: evidence, mechanisms, and recommendations. Ann N Y Acad Sci 2012 Oct;1271(1):37-43 [FREE Full text] [doi: 10.1111/j.1749-6632.2012.06750.x] [Medline: 23050962]
5. Butryn ML, Webb V, Wadden TA. Behavioral treatment of obesity. Psychiatr Clin North Am 2011 Dec;34(4):841-859 [FREE Full text] [doi: 10.1016/j.psc.2011.08.006] [Medline: 22098808]
6. Jensen MD, Ryan DH, Apovian CM, Ard JD, Comuzzie AG, Donato KA, American College of Cardiology/American Heart Association Task Force on Practice Guidelines, Obesity Society. 2013 AHA/ACC/TOS guideline for the management of overweight and obesity in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and The Obesity Society. J Am Coll Cardiol 2014 Jul 01;63(25 Pt B):2985-3023 [FREE Full text] [doi: 10.1016/j.jacc.2013.11.004] [Medline: 24239920]
7. Wadden TA, Tronieri JS, Butryn ML. Lifestyle modification approaches for the treatment of obesity in adults. Am Psychol 2020;75(2):235-251 [FREE Full text] [doi: 10.1037/amp0000517] [Medline: 32052997]
8. Butryn ML, Godfrey KM, Martinelli MK, Roberts SR, Forman EM, Zhang F. Digital self-monitoring: does adherence or association with outcomes differ by self-monitoring target? Obes Sci Pract 2019 Dec 12;6(2):126-133 [FREE Full text] [doi: 10.1002/osp4.391] [Medline: 32313670]
9. Burke LE, Wang J, Sevick MA. Self-monitoring in weight loss: a systematic review of the literature. J Am Diet Assoc 2011 Jan;111(1):92-102 [FREE Full text] [doi: 10.1016/j.jada.2010.10.008] [Medline: 21185970]
10. Tronieri JS, Wadden TA, Walsh O, Berkowitz RI, Alamuddin N, Chao AM. Measures of adherence as predictors of early and total weight loss with intensive behavioral therapy for obesity combined with liraglutide 3.0 mg. Behav Res Ther 2020 Aug;131:103639 [FREE Full text] [doi: 10.1016/j.brat.2020.103639] [Medline: 32450367]
11. Chen J, Berkman W, Bardouh M, Ng CY, Allman-Farinelli M. The use of a food logging app in the naturalistic setting fails to provide accurate measurements of nutrients and poses usability challenges. Nutrition 2019 Jan;57:208-216. [doi: 10.1016/j.nut.2018.05.003] [Medline: 30184514]
12. Solbrig L, Jones R, Kavanagh D, May J, Parkin T, Andrade J. People trying to lose weight dislike calorie counting apps and want motivational support to help them achieve their goals. Internet Interv 2017 Mar;7:23-31 [FREE Full text] [doi: 10.1016/j.invent.2016.12.003] [Medline: 28286739]
13. Burke LE, Sereika SM, Music E, Warziski M, Styn MA, Stone A. Using instrumented paper diaries to document self-monitoring patterns in weight loss. Contemp Clin Trials 2008 Mar;29(2):182-193 [FREE Full text] [doi: 10.1016/j.cct.2007.07.004] [Medline: 17702667]
14. Peterson ND, Middleton KR, Nackers LM, Medina KE, Milsom VA, Perri MG. Dietary self-monitoring and long-term success with weight management. Obesity (Silver Spring) 2014 Sep;22(9):1962-1967 [FREE Full text] [doi: 10.1002/oby.20807] [Medline: 24931055]
15. Turner-McGrievy GM, Dunn CG, Wilcox S, Boutté AK, Hutto B, Hoover A, et al. Defining adherence to mobile dietary self-monitoring and assessing tracking over time: tracking at least two eating occasions per day is best marker of adherence within two different mobile health randomized weight loss interventions. J Acad Nutr Diet 2019 Sep;119(9):1516-1524 [FREE Full text] [doi: 10.1016/j.jand.2019.03.012] [Medline: 31155473]
16. Evert AB, Franz MJ. Why weight loss maintenance is difficult. Diabetes Spectr 2017 Aug;30(3):153-156 [FREE Full text] [doi: 10.2337/ds017-0025] [Medline: 28848306]
17. Maclean PS, Bergouignan A, Cornier MA, Jackman MR. Biology's response to dieting: the impetus for weight regain. Am J Physiol Regul Integr Comp Physiol 2011 Sep;301(3):R581-R600 [FREE Full text] [doi: 10.1152/ajpregu.00755.2010] [Medline: 21677272]

18. Hall KD, Kahan S. Maintenance of lost weight and long-term management of obesity. Med Clin North Am 2018 Jan;102(1):183-197 [FREE Full text] [doi: 10.1016/j.mcna.2017.08.012] [Medline: 29156185]
19. Dinu M, Abbate R, Gensini GF, Casini A, Sofi F. Vegetarian, vegan diets and multiple health outcomes: a systematic review with meta-analysis of observational studies. Crit Rev Food Sci Nutr 2017 Nov 22;57(17):3640-3649. [doi: 10.1080/10408398.2016.1138447] [Medline: 26853923]
20. Huang RY, Huang CC, Hu FB, Chavarro JE. Vegetarian diets and weight reduction: a meta-analysis of randomized controlled trials. J Gen Intern Med 2016 Jan;31(1):109-116 [FREE Full text] [doi: 10.1007/s11606-015-3390-7] [Medline: 26138004]
21. Kahleova H, Levin S, Barnard ND. Vegetarian dietary patterns and cardiovascular disease. Prog Cardiovasc Dis 2018;61(1):54-61. [doi: 10.1016/j.pcad.2018.05.002] [Medline: 29800598]
22. Toumpanakis A, Turnbull T, Alba-Barba I. Effectiveness of plant-based diets in promoting well-being in the management of type 2 diabetes: a systematic review. BMJ Open Diabetes Res Care 2018 Oct 30;6(1):e000534 [FREE Full text] [doi: 10.1136/bmjdrc-2018-000534] [Medline: 30487971]
23. Ello-Martin JA, Roe LS, Ledikwe JH, Beach AM, Rolls BJ. Dietary energy density in the treatment of obesity: a year-long trial comparing 2 weight-loss diets. Am J Clin Nutr 2007 Jun;85(6):1465-1477 [FREE Full text] [doi: 10.1093/ajcn/85.6.1465] [Medline: 17556681]
24. Hall KD, Guo J, Courville AB, Boring J, Brychta R, Chen KY, et al. Effect of a plant-based, low-fat diet versus an animal-based, ketogenic diet on ad libitum energy intake. Nat Med 2021 Feb;27(2):344-353. [doi: 10.1038/s41591-020-01209-1] [Medline: 33479499]
25. Rolls BJ, Drewnowski A, Ledikwe JH. Changing the energy density of the diet as a strategy for weight management. J Am Diet Assoc 2005 May;105(5 Suppl 1):S98-103. [doi: 10.1016/j.jada.2005.02.033] [Medline: 15867904]
26. Warrilow A, Mellor D, McKune A, Pumpa K. Dietary fat, fibre, satiation, and satiety-a systematic review of acute studies. Eur J Clin Nutr 2019 Mar;73(3):333-344. [doi: 10.1038/s41430-018-0295-7] [Medline: 30166637]
27. Rolls BJ, Castellanos VH, Halford JC, Kilara A, Panyam D, Pelkman CL, et al. Volume of food consumed affects satiety in men. Am J Clin Nutr 1998 Jun;67(6):1170-1177. [doi: 10.1093/ajcn/67.6.1170] [Medline: 9625090]
28. Champagne CM, Bray G, Sacks F, Miketinas D. Fiber intake, dietary energy density, and diet-type predict 6-month weight-loss in free-living adults who adhered to prescribed macronutrient and energy composition of varying diets. FASEB J 2017 Apr;31(S1):796.3. [doi: 10.1096/fasebj.31.1_supplement.796.3]
29. Crowe FL, Appleby PN, Travis RC, Key TJ. Risk of hospitalization or death from ischemic heart disease among British vegetarians and nonvegetarians: results from the EPIC-Oxford cohort study. Am J Clin Nutr 2013 Mar;97(3):597-603. [doi: 10.3945/ajcn.112.044073] [Medline: 23364007]
30. Ferdowsian HR, Barnard ND, Hoover VJ, Katcher HI, Levin SM, Green AA, et al. A multicomponent intervention reduces body weight and cardiovascular risk at a GEICO corporate site. Am J Health Promot 2010;24(6):384-387. [doi: 10.4278/ajhp.081027-QUAN-255] [Medline: 20594095]
31. Yokoyama Y, Barnard ND, Levin SM, Watanabe M. Vegetarian diets and glycemic control in diabetes: a systematic review and meta-analysis. Cardiovasc Diagn Ther 2014 Oct;4(5):373-382 [FREE Full text] [doi: 10.3978/j.issn.2223-3652.2014.10.04] [Medline: 25414824]
32. Madigan M, Karhu E. The role of plant-based nutrition in cancer prevention. J Unexplored Med Data 2018 Nov 08;3:9. [doi: 10.20517/2572-8180.2018.05]
33. Poore J, Nemecek T. Reducing food's environmental impacts through producers and consumers. Science 2018 Jun 01;360(6392):987-992. [doi: 10.1126/science.aaq0216] [Medline: 29853680]
34. Turner-McGrievy GM, Davidson CR, Wingard EE, Billings DL. Low glycemic index vegan or low-calorie weight loss diets for women with polycystic ovary syndrome: a randomized controlled feasibility study. Nutr Res 2014 Jun;34(6):552-558. [doi: 10.1016/j.nutres.2014.04.011] [Medline: 25026923]
35. Dall T, Chakrabarti R, Iacobucci W, Hansari A, West T. The complexities of physician supply and demand: projections from 2015 to 2030. Association of American Medical Colleges. 2017. URL: https://www.luminafoundation.org/files/resources/aamc-projections-update-2017.pdf [accessed 2021-06-01]
36. Rollo ME, Burrows T, Vincze LJ, Harvey J, Collins CE, Hutchesson MJ. Cost evaluation of providing evidence-based dietetic services for weight management in adults: in-person versus eHealth delivery. Nutr Diet 2018 Feb;75(1):35-43. [doi: 10.1111/1747-0080.12335] [Medline: 29411491]
37. Bradley LE, Smith-Mason CE, Corsica JA, Kelly MC, Hood MM. Remotely delivered interventions for obesity treatment. Curr Obes Rep 2019 Dec;8(4):354-362. [doi: 10.1007/s13679-019-00356-y] [Medline: 31701350]
38. Hutchesson MJ, Tan CY, Morgan P, Callister R, Collins C. Enhancement of self-monitoring in a Web-based weight loss program by extra individualized feedback and reminders: randomized trial. J Med Internet Res 2016 Apr 12;18(4):e82 [FREE Full text] [doi: 10.2196/jmir.4100] [Medline: 27072817]
39. Thomas JG, Leahey TM, Wing RR. An automated internet behavioral weight-loss program by physician referral: a randomized controlled trial. Diabetes Care 2015 Jan;38(1):9-15 [FREE Full text] [doi: 10.2337/dc14-1474] [Medline: 25404659]
40. Turner-McGrievy G, Mandes T, Crimarco A. A plant-based diet for overweight and obesity prevention and treatment. J Geriatr Cardiol 2017 May;14(5):369-374 [FREE Full text] [doi: 10.11909/j.issn.1671-5411.2017.05.002] [Medline: 28630616]

41. Spahn JM, Reeves RS, Keim KS, Laquatra I, Kellogg M, Jortberg B, et al. State of the evidence regarding behavior change theories and strategies in nutrition counseling to facilitate health and food behavior change. J Am Diet Assoc 2010 Jun;110(6):879-891. [doi: 10.1016/j.jada.2010.03.021] [Medline: 20497777]
42. Haggerty AF, Hagemann A, Barnett M, Thornquist M, Neuhouser ML, Horowitz N, et al. A randomized, controlled, multicenter study of technology-based weight loss interventions among endometrial cancer survivors. Obesity (Silver Spring) 2017 Nov;25 Suppl 2:S102-S108 [FREE Full text] [doi: 10.1002/oby.22021] [Medline: 29086522]
43. Bowen DJ, Kreuter M, Spring B, Cofta-Woerpel L, Linnan L, Weiner D, et al. How we design feasibility studies. Am J Prev Med 2009 May;36(5):452-457 [FREE Full text] [doi: 10.1016/j.amepre.2009.02.002] [Medline: 19362699]
44. Bradley LE, Forman EM, Kerrigan SG, Goldstein SP, Butryn ML, Thomas JG, et al. Project HELP: a remotely delivered behavioral intervention for weight regain after bariatric surgery. Obes Surg 2017 Mar;27(3):586-598. [doi: 10.1007/s11695-016-2337-3] [Medline: 27586525]
45. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ 2014 Mar 07;348:g1687. [doi: 10.1136/bmj.g1687] [Medline: 24609605]
46. Eldridge SM, Lancaster GA, Campbell MJ, Thabane L, Hopewell S, Coleman CL, et al. Defining feasibility and pilot studies in preparation for randomised controlled trials: development of a conceptual framework. PLoS One 2016 Mar 15;11(3):e0150205 [FREE Full text] [doi: 10.1371/journal.pone.0150205] [Medline: 26978655]
47. Wright N, Wilson L, Smith M, Duncan B, McHugh P. The BROAD study: a randomised controlled trial using a whole food plant-based diet in the community for obesity, ischaemic heart disease or diabetes. Nutr Diabetes 2017 Mar 20;7(3):e256 [FREE Full text] [doi: 10.1038/nutd.2017.3] [Medline: 28319109]
48. Barnard ND, Gloede L, Cohen J, Jenkins DJ, Turner-McGrievy G, Green AA, et al. A low-fat vegan diet elicits greater macronutrient changes, but is comparable in adherence and acceptability, compared with a more conventional diabetes diet among individuals with type 2 diabetes. J Am Diet Assoc 2009 Feb;109(2):263-272 [FREE Full text] [doi: 10.1016/j.jada.2008.10.049] [Medline: 19167953]
49. Forman EM, Butryn ML. Effective Weight Loss: An Acceptance-Based Behavioral Approach, Clinician Guide. Oxford, UK: Oxford University Press; 2016.
50. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
51. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004 Apr;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]
52. Lisle DJ, Goldhamer A. The Pleasure Trap: Mastering the Force that Undermines Health & Happiness. Summertown, TN, USA: Book Publishing Company; 2007.
53. The Latest Nutrition Related Topics. NutritionFacts. URL: https://nutritionfacts.org/ [accessed 2020-07-18]
54. McDougall J, McDougall M. The Starch Solution: Eat the Foods You Love, Regain Your Health, and Lose the Weight for Good!. Emmaus, PA, USA: Rodale Books; 2013.
55. Diabetes Prevention Program (DPP) Research Group. The Diabetes Prevention Program (DPP): description of lifestyle intervention. Diabetes Care 2002 Dec;25(12):2165-2171 [FREE Full text] [doi: 10.2337/diacare.25.12.2165] [Medline: 12453955]
56. Forman EM, Butryn ML, Juarascio AS, Bradley LE, Lowe MR, Herbert JD, et al. The mind your health project: a randomized controlled trial of an innovative behavioral treatment for obesity. Obesity (Silver Spring) 2013 Jun;21(6):1119-1126 [FREE Full text] [doi: 10.1002/oby.20169] [Medline: 23666772]
57. Duhigg C. The Power of Habit: Why We Do What We Do in Life and Business. New York, NY, USA: Random House Books; 2012.
58. Marlatt GA, Donovan DM. Relapse Prevention: Maintenance Strategies in the Treatment of Addictive Behaviors. 2nd edition. New York, NY, USA: The Guilford Press; 2005.
59. Hettema J, Steele J, Miller WR. Motivational interviewing. Annu Rev Clin Psychol 2005;1:91-111. [doi: 10.1146/annurev.clinpsy.1.102803.143833] [Medline: 17716083]
60. Zheng Y, Klem ML, Sereika SM, Danford CA, Ewing LJ, Burke LE. Self-weighing in weight management: a systematic literature review. Obesity (Silver Spring) 2015 Feb;23(2):256-265 [FREE Full text] [doi: 10.1002/oby.20946] [Medline: 25521523]
61. Krukowski RA, Ross KM. Measuring weight with electronic scales in clinical and research settings during the coronavirus disease 2019 pandemic. Obesity (Silver Spring) 2020 Jul;28(7):1182-1183 [FREE Full text] [doi: 10.1002/oby.22851] [Medline: 32339394]
62. Ross KM, Wing RR. Concordance of in-home "smart" scale measurement with body weight measured in-person. Obes Sci Pract 2016 Jun;2(2):224-248 [FREE Full text] [doi: 10.1002/osp4.41] [Medline: 27499884]
63. Chambliss HO, Huber RC, Finley CE, McDoniel SO, Kitzman-Ulrich H, Wilkinson WJ. Computerized self-monitoring and technology-assisted feedback for weight loss with and without an enhanced behavioral component. Patient Educ Couns 2011 Dec;85(3):375-382. [doi: 10.1016/j.pec.2010.12.024] [Medline: 21295433]

64. Joiner KL, Nam S, Whittemore R. Lifestyle interventions based on the diabetes prevention program delivered via eHealth: a systematic review and meta-analysis. Prev Med 2017 Jul;100:194-207 [FREE Full text] [doi: 10.1016/j.ypmed.2017.04.033] [Medline: 28456513]
65. Williamson DA, Bray GA, Ryan DH. Is 5% weight loss a satisfactory criterion to define clinically significant weight loss? Obesity (Silver Spring) 2015 Dec;23(12):2319-2320 [FREE Full text] [doi: 10.1002/oby.21358] [Medline: 26523739]
66. Clarys P, Deliens T, Huybrechts I, Deriemaeker P, Vanaelst B, De Keyzer W, et al. Comparison of nutritional quality of the vegan, vegetarian, semi-vegetarian, pesco-vegetarian and omnivorous diet. Nutrients 2014 Mar 24;6(3):1318-1332 [FREE Full text] [doi: 10.3390/nu6031318] [Medline: 24667136]
67. Ware Jr JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992 Jun;30(6):473-483. [Medline: 1593914]
68. Kolotkin RL, Meter K, Williams GR. Quality of life and obesity. Obes Rev 2001 Nov;2(4):219-229. [doi: 10.1046/j.1467-789x.2001.00040.x] [Medline: 12119993]
69. Ferguson RJ, Robinson AB, Splaine M. Use of the reliable change index to evaluate clinical significance in SF-36 outcomes. Qual Life Res 2002 Sep;11(6):509-516. [doi: 10.1023/a:1016350431190] [Medline: 12206571]
70. Batterham MJ, Tapsell LC, Charlton KE. Analyzing weight loss intervention studies with missing data: which methods should be used? Nutrition 2013;29(7-8):1024-1029. [doi: 10.1016/j.nut.2013.01.017] [Medline: 23644010]
71. Krukowski RA, Tilford JM, Harvey-Berino J, West DS. Comparing behavioral weight loss modalities: incremental cost-effectiveness of an Internet-based versus an in-person condition. Obesity (Silver Spring) 2011 Aug;19(8):1629-1635 [FREE Full text] [doi: 10.1038/oby.2010.341] [Medline: 21253001]
72. Forman-Hoffman V, Little A, Wahls T. Barriers to obesity management: a pilot study of primary care clinicians. BMC Fam Pract 2006 Jun 06;7:35 [FREE Full text] [doi: 10.1186/1471-2296-7-35] [Medline: 16756673]
73. Dansinger ML, Gleason JA, Griffith JL, Selker HP, Schaefer EJ. Comparison of the Atkins, Ornish, Weight Watchers, and Zone diets for weight loss and heart disease risk reduction: a randomized trial. JAMA 2005 Jan 05;293(1):43-53. [doi: 10.1001/jama.293.1.43] [Medline: 15632335]
74. Spring B, Pfammatter AF, Marchese SH, Stump T, Pellegrini C, McFadden HG, et al. A factorial experiment to optimize remotely delivered behavioral treatment for obesity: results of the Opt-IN study. Obesity (Silver Spring) 2020 Sep;28(9):1652-1662 [FREE Full text] [doi: 10.1002/oby.22915] [Medline: 32656994]
75. Burke LE, Swigart V, Warziski Turk M, Derro N, Ewing LJ. Experiences of self-monitoring: successes and struggles during treatment for weight loss. Qual Health Res 2009 Jun;19(6):815-828 [FREE Full text] [doi: 10.1177/1049732309335395] [Medline: 19365099]
76. Forman EM, Goldstein SP, Zhang F, Evans BC, Manasse SM, Butryn ML, et al. OnTrack: development and feasibility of a smartphone app designed to predict and prevent dietary lapses. Transl Behav Med 2019 Mar 01;9(2):236-245 [FREE Full text] [doi: 10.1093/tbm/iby016] [Medline: 29617911]
77. Benton D, Young HA. Reducing calorie intake may not help you lose body weight. Perspect Psychol Sci 2017 Sep;12(5):703-714 [FREE Full text] [doi: 10.1177/1745691617690878] [Medline: 28657838]
78. U.S. Census Bureau Releases New Educational Attainment Data. United States Census Bureau. 2020 Mar 30. URL: https://www.census.gov/newsroom/press-releases/2020/educational-attainment.html [accessed 2021-06-28]
79. Sherwood NE, Morton N, Jeffery RW, French SA, Neumark-Sztainer D, Falkner NH. Consumer preferences in format and type of community-based weight control programs. Am J Health Promot 1998;13(1):12-18. [doi: 10.4278/0890-1171-13.1.12] [Medline: 10186930]

Abbreviations

BCT: behavioral change technique
CONSORT: Consolidated Standards of Reporting Trials
ITT: intention-to-treat
PP: per-protocol
S-BT: standard behavioral weight loss treatment
SF-36: 36-item Short-Form General Health Survey
WELL Center: Drexel University Center for Weight, Eating and Lifestyle Science
WFPBD: whole food plant-based diet

Edited by A Mavragani; submitted 26.02.22; peer-reviewed by J Andrade, P Grace-Farfaglia; comments to author 06.04.22; revised version received 23.05.22; accepted 24.05.22; published 23.06.22.

Please cite as:
Chwyl C, Wright N, M Turner-McGrievy G, L Butryn M, M Forman E
Remotely Delivered Behavioral Weight Loss Intervention Using an Ad Libitum Plant-Based Diet: Pilot Acceptability, Feasibility, and Preliminary Results
JMIR Form Res 2022;6(6):e37414
URL: https://formative.jmir.org/2022/6/e37414
doi:10.2196/37414
PMID:35737443

Original Paper

Identifying Patients With Delirium Based on Unstructured Clinical Notes: Observational Study

Wendong Ge[1], PhD; Haitham Alabsi[1], MD; Aayushee Jain[1], MSc; Elissa Ye[1], MSc; Haoqi Sun[1], PhD; Marta Fernandes[1], PhD; Colin Magdamo[1], BS; Ryan A Tesh[1], BS; Sarah I Collens[1], BS; Amy Newhouse[1], MD; Lidia MVR Moura[1], MD, PhD, MPH; Sahar Zafar[1], MD; John Hsu[1], MD; Oluwaseun Akeju[1], MD; Gregory K Robbins[1], MD; Shibani S Mukerji[1], MD, PhD; Sudeshna Das[1], PhD; M Brandon Westover[1], MD, PhD

Massachusetts General Hospital, Boston, MA, United States

Corresponding Author:
M Brandon Westover, MD, PhD
Massachusetts General Hospital
50 Staniford Street
Boston, MA, 02114
United States
Phone: 1 650 862 1154
Email: mwestover@mgh.harvard.edu

Abstract

Background: Delirium in hospitalized patients is a syndrome of acute brain dysfunction. Diagnostic (International Classification of Diseases [ICD]) codes are often used in studies using electronic health records (EHRs), but they are inaccurate.

Objective: We sought to develop a more accurate method using natural language processing (NLP) to detect delirium episodes on the basis of unstructured clinical notes.

Methods: We collected 1.5 million notes from >10,000 patients from among 9 hospitals. Seven experts iteratively labeled 200,471 sentences. Using these, we trained three NLP classifiers: Support Vector Machine, Recurrent Neural Networks, and Transformer. Testing was performed using an external data set. We also evaluated associations with delirium billing (ICD) codes, medications, orders for restraints and sitters, direct assessments (Confusion Assessment Method [CAM] scores), and in-hospital mortality. F1 scores, confusion matrices, and areas under the receiver operating characteristic curve (AUCs) were used to compare NLP models. We used the ϕ coefficient to measure associations with other delirium indicators.

Results: The transformer NLP performed best on the following parameters: micro F1=0.978, macro F1=0.918, positive AUC=0.984, and negative AUC=0.992. NLP detections exhibited higher correlations (ϕ) than ICD codes with deliriogenic medications (0.194 vs 0.073 for ICD codes), restraints and sitter orders (0.358 vs 0.177), mortality (0.216 vs 0.000), and CAM scores (0.256 vs –0.028).

Conclusions: Clinical notes are an attractive alternative to ICD codes for EHR delirium studies but require automated methods. Our NLP model detects delirium with high accuracy, similar to manual chart review. Our NLP approach can provide more accurate determination of delirium for large-scale EHR-based studies regarding delirium, quality improvement, and clinical trails.

(JMIR Form Res 2022;6(6):e33834) doi:10.2196/33834

KEYWORDS

delirium; electronic health records; clinical notes; machine learning; natural language processing

Introduction

Delirium is an acute neuropsychiatric syndrome with features of inattention and global cognitive dysfunction, associated with increased hospital length of stay, in-hospital mortality, and long-term cognitive disability [1]. Delirium occurs in up to 26% of hospitalized patients; prevalence rates may reach 42% in patients older than age 65 years [2].

Electronic health records (EHRs) offer a rich source of information for studies of delirium; however, determining which patients have delirium is challenging. Manual review of medical records is time consuming, limiting studies to a small fraction of patients at risk. A more scalable approach is to use International Classification of Diseases (ICD) billing codes. This approach was recently used by a study [3] to assess 200 patients admitted to a skilled nursing facility, revealing that

ICD codes achieved 96.0% specificity but only 53.1% sensitivity. Another study [4] analyzed clinical data from 184 older adults at one academic medical center and found that ICD codes had a specificity of 98% and sensitivity of 18%. Thus, ICD codes miss a large fraction of patients with delirium.

On the other hand, rich information about patients' status exists in narrative clinical notes from doctors, nurses, physical therapists, and other health care workers [5]. However, extracting this information is challenging because of the flexibility of natural language.

In this work, we collected 1.5 million clinical notes from over 10,000 patients from 7 distinct cohorts from among 9 hospitals and developed a natural language processing (NLP) algorithm to identify patients with delirium from unstructured EHR notes.

Methods

Data Set Description and Sentence Extraction

We collected 1,565,678 clinical notes from 10,516 patients from 9 hospitals, including Massachusetts General Hospital, Brigham and Women's Hospital, Cooley Dickinson Hospital, Martha's Vineyard Hospital, McLean Hospital, Nantucket Cottage Hospital, Newton-Wellesley Hospital, North Shore Medical Center, and Spaulding Rehabilitation Hospital. These 10,516 patients were from 7 previously assembled cohort studies:

- Antiepileptic drug (AED) data set: this data set comprises patients who received AEDs and is used to study adverse effects of AEDs (n=852).
- GIFTS data set: this data set comprises older patients admitted for orthopedic surgery and is used to study delirium (n=576).
- Dementia data set: this data set comprises patients who were at risk for dementia and is used to study dementia (n=802).
- COVID-19 data set: this data set comprises patients who were hospitalized for COVID-19 and is used to study hospitalization, intensive care unit admission, intubation, and mortality prediction for patients with COVID-19 (n=3429).
- NCC data set: this data set is used to study neurological diseases such as delirium, headache, and anosmia for patients at neurocritical care units (n=1985).
- LTM data set: this data set comprises acutely ill patients undergoing continuous electroencephalographic monitoring (n=395). These patients underwent in-person delirium assessments by research staff. Thus, this data set contains assessment records rather than clinical notes.
- Control data set: this data set comprises inpatients randomly selected as a control group from the Massachusetts General Brigham hospital system (n=2477).

Demographic features of these cohorts are shown in Multimedia Appendix 1.

Creating the Gold Standard: Sentence Labeling

We first created a comprehensive collection of keywords related to delirium; these included the following: "delirium," "delirious," "encephalopathy," "confused," "confusion," "agitated," "agitation," "inattentive," "inattention," "disorient," "disoriented," "disorientation," "reorient," "restraints," "lethargy," "psychosis," "hallucination," "inappropriate behavior," "fluctuating arousal," "altered mental status," "mental status change," "fluctuating mental status," and "waxing and waning mental status." We extracted all sentences containing any of these keywords from the assembled collection of notes.

Next, we created a gold-standard set of labels for sentences. Examples are shown in Multimedia Appendix 2.

We developed a graphical user interface (GUI) for efficient iterative labeling of sentences. Active learning, an algorithm to select the most informative samples, was used to select candidate sentences in each round. The labeling process was as follows:

- Step 0: candidate sentences were randomly selected from the set of unlabeled sentences.
- Step 1: experts labeled candidate sentences and created regular expressions called "always patterns" (described below in Regular Expression Generation).
- Step 2: unlabeled sentences were screened for "always patterns," corresponding labels were assigned to sentences that match, and these were added to the labeled set.
- Step 3: the labeled sentences were used to train a classifier (introduced in Prediction Model).
- Step 4: the classifier was used to scan unlabeled sentences and assign them a label and an embedding vector.
- Step 5: sentence embedding vectors were used to generate an embedding map via Uniform Manifold Approximation and Projection [6].
- Step 6: candidate sentences were selected from the unlabeled data set with two query strategies: uncertainty based on the entropy of prediction scores and diversity based on the embedding map (Multimedia Appendix 3). Each query selects half of the candidate sentences for the next round. Then, the process was reverted to step 1.

Regular Expression Generation

While labeling sentences, experts created "always patterns": a regular expression that, when present, warrants assigning the corresponding label to the sentence. Multimedia Appendix 2 provides examples of "always patterns" for positive, negative, and neither patterns. The GUI used "always patterns" to scan the residual unlabeled sentences to assign a label to all matched sentences, thus enhancing labeling efficiency.

Prediction Model

We developed three models to identify delirium sentences: Support Vector Machine (SVM), long short-term memory (LSTM), and Transformer models. The LSTM model was also used in active learning when collecting labels. Details of the three models are as follows.

SVM is a widely used text classifier based on a "bag of words" representation [7]. Sentences with delirium-related keywords are first transformed into sentence vectors via "a bag of unigrams and bigrams," and the SVM algorithm finds hyperplanes that separate different categories. The distances between sample points and hyperplanes are used to calculate prediction scores.

Recurrent neural networks with LSTM units (RNN-LSTM) are common models for sequence learning, where an LSTM unit contains a cell for memory, an input gate to control input information flow, an output gate to control output information flow, and a forgetting gate to update memory [8]. We used a 3-layer bidirectional RNN with LSTM units to encode sentences. The vector representation corresponding to the keyword location was used for classification.

A transformer is a previously proposed [9] transduction model that computes a representation of each word in a sentence relying on self-attention. It is also the model used in Bidirectional Encoder Representations from Transformers (BERT) [10]. We used a 3-layer Transformer model to transform a sentence into a sequence of vectors. The vector representation corresponding to the delirium keyword was then used for classification. The word vectors from BERT were used as initial vectors.

Comparison of Delirium NLP Results With Other Delirium Indicators

To evaluate construct validity of our EHR-based delirium detection algorithms, we evaluated the strength of the association between presence of delirium as detected by our NLP models with other clinical outcomes or events known to be associated with delirium. These included the use of ICD billing codes for delirium; use of medications related to delirium; use of restraints and sitters; and in-hospital mortality. For one cohort (the LTM data set) we had access to one-time in-person delirium assessments using the Confusion Assessment Method (CAM), which has been already been validated as a good proxy for DSM-5 in prior studies. For these, we compared the presence of delirium, as defined by CAM, with the presence of positive delirium sentences in clinical notes during hospitalization. Details are provided in Multimedia Appendix 4.

Interrater Agreement

Pairwise interrater agreement (IRA) is used to measure agreement between human and human (model) for each category. Details are provided in Multimedia Appendix 5.

Data Split for Evaluation

We combined the AED, GIFTS, Dementia, COVID-19, NCC, and Control data sets to yield a data set for sentence labeling based on active learning. We collected 200,471 labeled sentences, including those directly labeled by human experts and those matched by “always patterns.” Of the 200,471 labeled sentences, 176,800 were “positive,” 15,577 were “negative,” and 8094 were “neither” sentences.

We designed two types of tests for NLP delirium detection algorithms: an internal test and an external test (see Multimedia Appendix 6).

Internal Test

In the internal test, we followed the standard machine learning evaluation pipeline, randomly splitting the 200,471 labeled sentences into a training data set (120,283 sentences, 60%), validation data set (40,094 sentences, 20%) for hyperparameter tuning, and test data set (40,094 sentences, 20%) for performance evaluation.

External Test

The LTM data set was not used for training the NLP algorithms. It was used entirely for testing. The LTM data set contained 16,067 sentences: 14,378 positive, 1193 negative, and 496 neither sentences.

Data Security and Ethics Approval

We have ethics approval (2013P001024) from the MassGeneral Brigham institutional review board to work with identified data internally. We will deidentify the data for sharing them with external partners to test and improve the models together. Some existing deidentification algorithms have been developed, such as the Phsyionet algorithm [11] and the Philter algorithm [12], but the recall of these algorithms is close to 100% rather than 100% perfect. Another option is federated learning, namely training the model across multiple decentralized machines holding local data by us and our external partners, without exchanging them.

Results

Performances of Delirium NLP classifiers

In the following analysis, the 95% CIs were calculated through bootstrapping [13].

Table 1 compares performances of SVM, RNN-LSTM, and Transformer on both internal and external tests. As the data set is an imbalanced multiclass data set, micro F1 scores, and macro F1 scores were used to evaluate performance [14]. When using micro F1 scores, the performance of the SVM, RNN-LSTM, and Transformer models was close on both the internal and external test sets. However, when using macro F1 scores, which measure average performance across categories, on the internal test the Transformer (0.927, 95% CI 0.925-0.930) performed similarly to the RNN-LSTM (0.922, 95% CI 0.920-0.925), and both Transformer and RNN-LSTM outperformed the SVM (0.839, 95% CI 0.835-0.842). In the external test set, the Transformer (0.918, 95% CI 0.914-0.921) displayed the best performance, while the SVM (0.885, 95% CI 0.881-0.889) displayed slightly better performance than the RNN-LSTM (0.868, 95% CI 0.862-0.874). Overall, the Transformer was thus the best model based on both micro F1 and macro F1 metrics.

Figure 1 illustrates confusion matrices for the best Transformer, normalized by row to show recall (sensitivity), and by column to show precision (positive predictive value). For the Positive category, precision and recall on both the internal and external test were close to 0.99. For the Negative category, on the internal test, precision (0.916, 95% CI 0.911-0.920) was slightly higher than recall (0.893, 95% CI 0.889-0.897), while on the external test, recall (0.947, 95% CI 0.942-0.951) was much higher than precision (0.861, 95% CI 0.852-0.870). For the Neither category, on both internal and external tests, precision (0.916, 95% CI 0.909-0.923 vs 0.886, 95% CI 0.877-0.894) was better than recall (0.867, 95% CI 0.860-0.873 vs 0.848, 95% CI 0.836-0.859). In summary, performance on the Negative category was better than that on the Neither category, and performance on the Positive category was better still.

Figure 2 compares receiver operating characteristic (ROC) curves and areas under the ROC curve (AUCs) for the Positive, Negative, and Neither categories on both internal and external tests. On the internal test data, the Transformer (Positive: 0.981, 95% CI 0.980-0.983; Negative: 0.985, 95% CI 0.984-0.986; Neither: 0.974, 95% CI 0.971-0.976) and RNN-LSTM (Positive: 0.980, 95% CI 0.978-0.981; Negative: 0.982, 95% CI 0.981-0.983; Neither: 0.972, 95% CI 0.969-0.974) were close, and both were better than SVM (Positive: 0.962, 95% CI 0.961-0.964; Negative: 0.962, 95% CI 0.961-0.963; Neither: 0.966, 95% CI 0.963-0.968).

On the external test, for the Positive category, the Transformer (0.984, 95% CI 0.983-0.985) was the best, and the SVM (0.974, 95% CI 0.972-0.976) was better than the RNN-LSTM (0.970, 95% CI 0.966-0.972). For the Negative category, the Transformer (0.992, 95% CI 0.991-0.993) was the best, followed by RNN-LSTM (0.984, 95% CI 0.982-0.985), and then the SVM (0.979, 95% CI 0.977-0.981). For the Neither category, the SVM (0.984, 95% CI 0.982-0.986) was the best, followed by the Transformer (0.969, 95% CI 0.967-0.973) and the RNN-LSTM (0.952, 95% CI 0.949-0.955).

We conclude that overall, the Transformer model performed the best. Hereinafter, “NLP” refers to the Transformer model.

Table 1. F1 scores for the Support Vector Machine, recurrent neural networks with long short-term model, and the Transformer model.

Scores	Support Vector Machine, mean (95% CI)	Recurrent neural networks with long short-term model, mean (95% CI)	Transformer, mean (95% CI)
Micro F1			
Internal test	0.949 (0.948-0.951)	0.977 (0.976-0.978)	0.978 (0.977-0.979)
External test	0.964 (0.963-0.966)	0.967 (0.965-0.968)	0.978 (0.977-0.979)
Macro F1			
Internal test	0.839 (0.835-0.842)	0.922 (0.920-0.925)	0.927 (0.925-0.930)
External test	0.885 (0.881-0.889)	0.868 (0.862-0.874)	0.918 (0.914-0.921)

Figure 1. Precision, recall, and F1 scores for delirium classifiers.

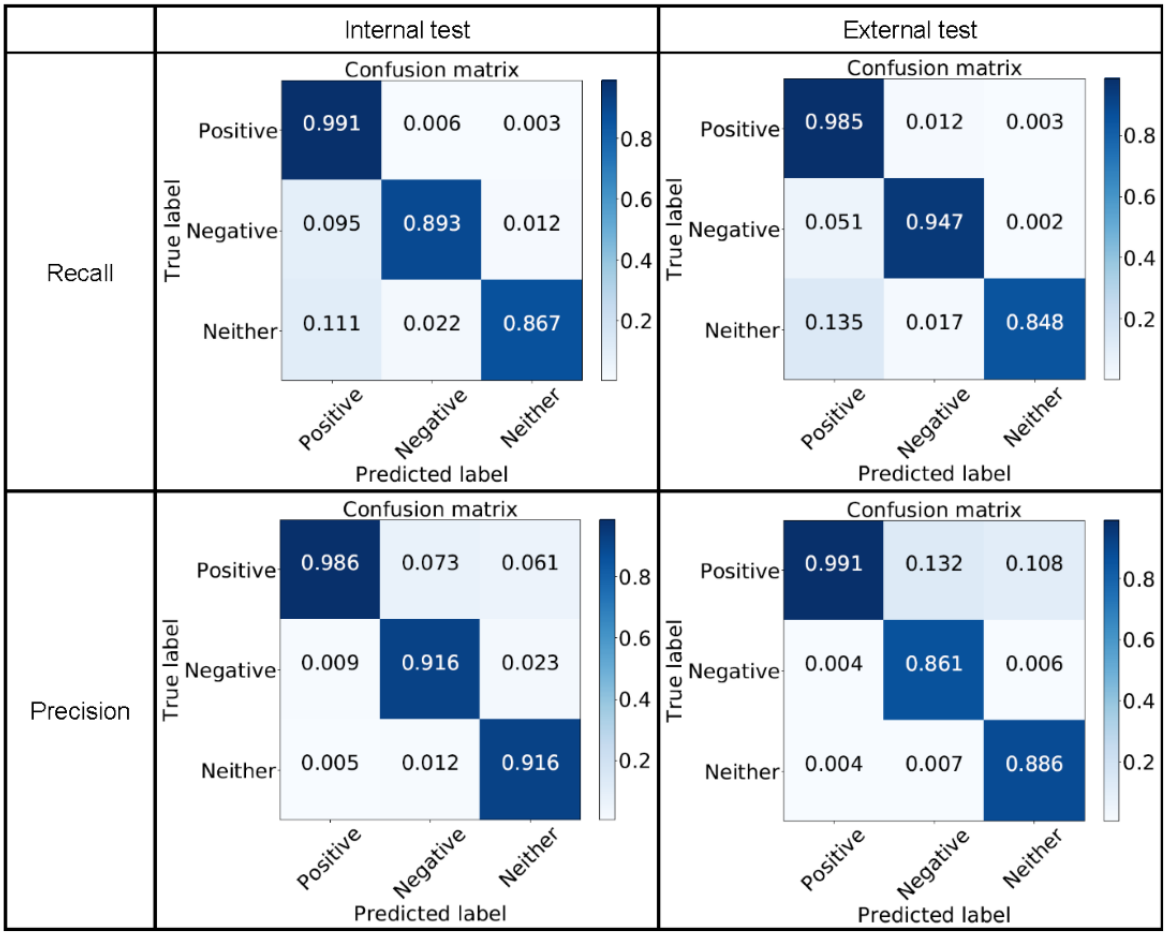
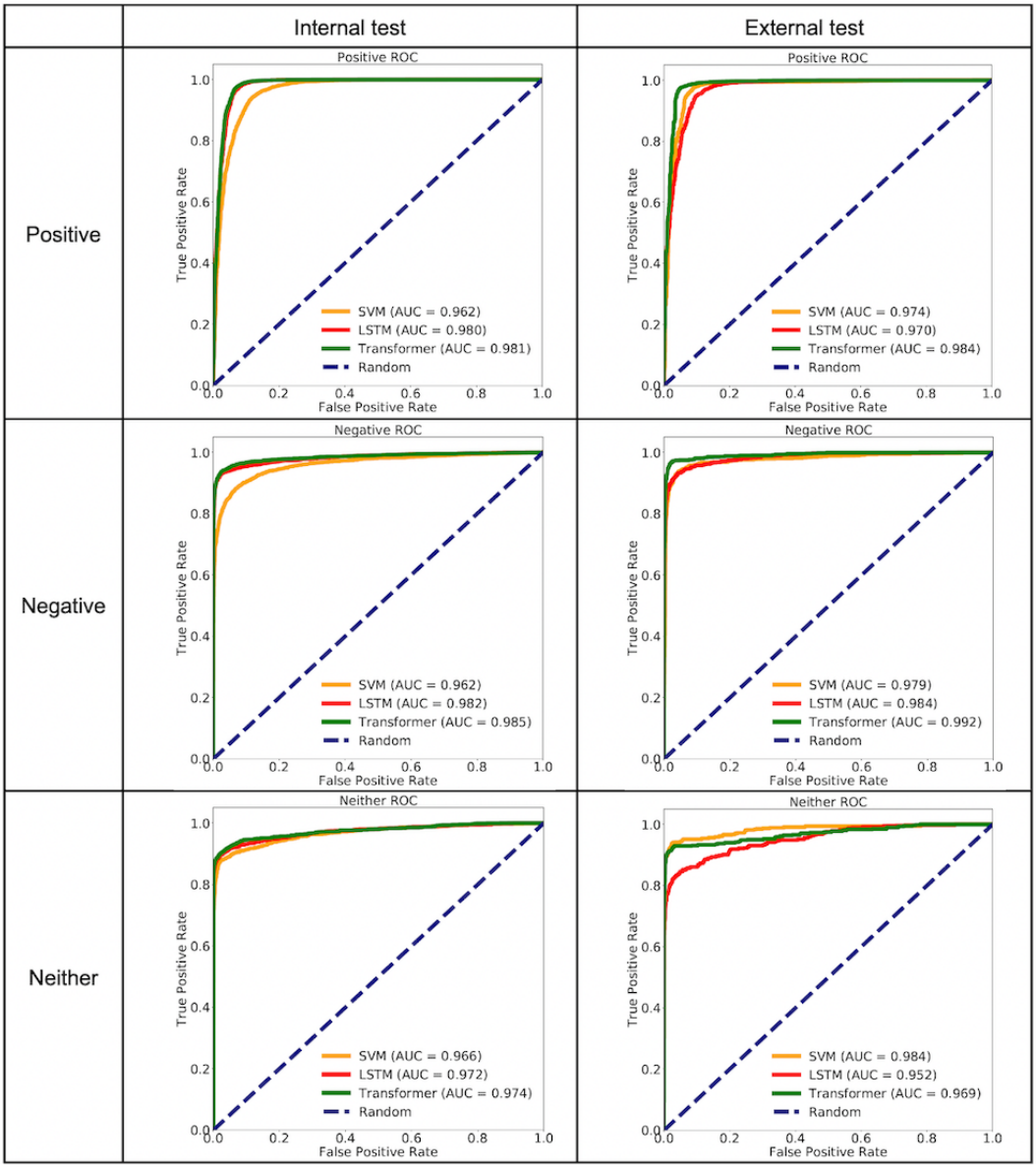


Figure 2. Receiver operating characteristic (ROC) curves for delirium classifiers. AUC: area under the curve, LSTM: long short-term model, SVM: Support Vector Machine.



Associations Between Delirium NLP Results and Other Delirium Indicators

Next, we compared associations between delirium NLP results and other delirium indicators. Results are shown in Table 2 For the NCC cohort (n=1985 patients), we assessed associations of NLP-detected delirium with delirium ICD code usage, medications, restraints and sitter orders, and mortality. For the LTM data set (n=395), we analyzed associations with CAM scores. For comparison, we also calculated the association of ICD code usage with the same delirium indicators.

We calculated these delirium indicators at the patient level, such that each patient is assigned a "+1" for NLP-based detection of delirium if they have one or more sentences classified as Positive by the NLP Transformer algorithm; otherwise, they were assigned a "–1." Similarly, patients were assigned scores of "+1" or "–1" for each of the other delirium indicators. We used the ϕ coefficient (mean square contingency coefficient) to measure associations between NLP-based delirium detections and each delirium indicator. When using our NLP detector to classify sentences in the NCC (or LTM) data set, the NCC (or LTM) data were only used as test data, as illustrated in Multimedia Appendix 6.

Table 2 shows that associations of delirium indicators with NLP results are much stronger than those with ICD codes.

In the NCC data set, the NLP model identified 1117 out of 1985 patients with positive delirium sentences (which were verified to be correct through manual review) but no delirium ICD codes. This highlights the low sensitivity of delirium ICD codes relative to manual chart review, and the excellent sensitivity of the NLP algorithm.

Table 2. Associations between delirium natural language processing indicators and other delirium indicators.

Data sets and delirium indicators	International Classification of Diseases codes, mean (95% CI)	Natural language processing classifiers, mean (95% CI)
NCC		
International Classification of Diseases codes	1	0.134 (0.133 to 0.135)
Medication	0.073 (0.072 to 0.074)	0.194 (0.192 to 0.197)
Restraints and sitter orders	0.177 (0.176 to 0.179)	0.358 (0.357 to 0.361)
Mortality	0.000 (–0.0002 to 0.0001)	0.216 (0.215 to 0.217)
LTM		
Confusion Assessment Method	–0.028 (–0.025 to –0.030)	0.256 (0.252 to 0.259)

Coverage Analysis

In creating the gold standard for labeling sentences, we developed many "always patterns" for delirium. While this set of sentences was large, we hypothesized that it might not be exhaustive; therefore, we investigated the coverage of our "always patterns" in another data set.

We analyzed the coverage of "always patterns" as follows. First, in the development data set (AED, GIFTS, Dementia, COVID-19, NCC, and control cohorts)—used for labeling the gold-standard set of sentences and for developing "always patterns"—97.6% (195,680) of sentences with delirium keywords were matched by at least one "always pattern." In the LTM data set, which was not used for labeling sentences, 78.2% (12,569) of sentences with delirium keywords matched at least one "always pattern."

We next tested the extent to which sentences not matched by "always patterns" were still accurately classified by the NLP model. To accomplish this, we randomly selected 400 sentences as follows:

- 100 sentences that both the Transformer and LSTM models predicted "Positive" for delirium
- 100 sentences that both the Transformer and LSTM models predicted "Negative" for delirium
- 100 sentences that both the Transformer and LSTM models predicted "Neither"; namely, not relevant to delirium
- 100 sentences on which the Transformer and LSTM models disagreed.

Two human experts (SM and MBW) independently labeled these 400 unmatched sentences. Pairwise IRA results are shown in Figure 3, where 95% CIs were calculated through Bootstrapping [13]. For unmatched sentences, the performance of model IRA (LSTM, Transformer) was close to that of human IRA for the Negative category but displayed gaps for Positive and Neither categories compared with human IRA.

We next investigated whether performance gaps in the new data set could be easily removed without repeating a large amount of sentence relabeling. For this investigation, we tried fine-tuning the Transformer model with a previously reported procedure [10]. This was readily done (green bars).

We conclude that the Transformer model is quite general, but not exhaustive; nevertheless, when gaps are encountered, the model can be readily tuned to accommodate previously unseen delirium sentence patterns.

Figure 4 illustrates mortality rates for the patients with different numbers of days with delirium in the GIFTS data set. The mortality rate increases monotonically with the number of delirium days.

Figure 3. Pairwise interrater agreement (IRA) for unmatched sentences. LSTM: long short-term memory.

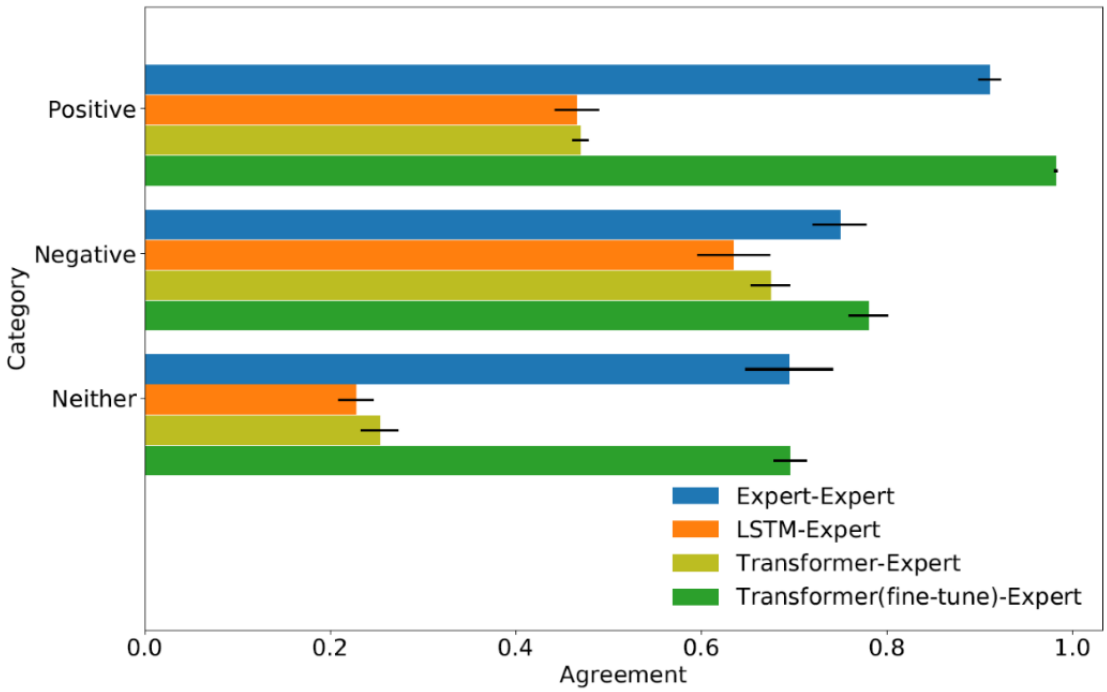
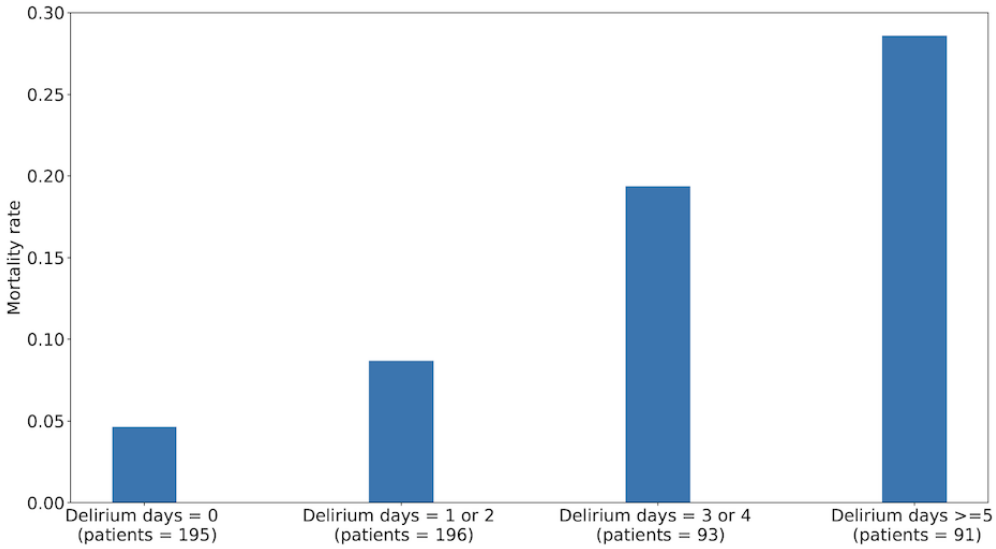


Figure 4. Mortality rate versus the number of days with delirium.



Discussion

Principal Findings

Our results show that an NLP approach can accurately detect patients with delirium, using unstructured clinical notes. These results are likely to be robust because they are based on a large collection of clinical notes from over 10,000 patients. The proposed delirium NLP approach is much more accurate, and especially more sensitive, than delirium ICD codes; it was able to detect patients who have delirium described in clinical notes but have no delirium ICD codes in their medical records. Further enhancing validity, NLP delirium detections are strongly associated with clinical factors known to be associated with delirium, including delirium-associated medications, use of restraints, and in-hospital mortality. This NLP tool will be useful for large-scale EHR research on delirium.

Application

The delirium NLP approach proposed in this work has many potential applications. First, the approach will be applied to many future large-scale studies regarding delirium, such as the causes of delirium and the effects of delirium on outcomes such as dementia. Second, the approach can review entire medical record in order to identify specific parts of the hospital, which seem to have more delirium, which can be used for quality

improvement. We can use this to identify factors (eg, medications) that might explain why delirium occurs. Third, the approach can be used to develop a delirium prediction model for clinical trials. The detection results of the NLP approach can be used as targets of prediction models, and the prediction models can be used to identify patients at a high risk for delirium, which provides information for interventions. The barriers of the applications are data and trust or transparency.

Comparison With Prior Work

Many prior studies have utilized ICD codes to identify delirium for large-scale EHR studies [3,4]. Our findings confirm observations from these earlier studies that ICD codes generally have high specificity but low sensitivity, leading to many missed cases of delirium. We investigated this finding in detail in the NCC cohort, where we observed that 1117 of 1985 patients who had positive delirium sentences had no corresponding delirium ICD codes. To confirm these findings, we used the NLP Transformer model to select the sentence with the highest positive score for each patient, and then manually reviewed the 1117 selected sentences, thereby manually confirming that these were true positives. These results show that the NLP approach largely overcomes the low sensitivity of delirium ICD codes.

NLP has been used to extract phenotypes from clinical notes in several previous studies. McCoy et al [15] used NLP to analyze discharge notes to improve prediction of suicide and accidental death after discharge. Gundlapalli et al [16] reported that a relatively simple case finding method based on string matching for specific keywords coupled with a negation algorithm and information extracted by a more complex NLP system could identify patients with inflammatory bowel disease. Zhou et al [17] applied an NLP approach to identify patients with depression on the basis of discharge summaries. Yang et al [18] explored transformer-based models for clinical concept extraction. Mascio et al [19] analyzed the impact of various word representations, text preprocessing, and classification algorithms on the performance of different text classification tasks based on EHRs. Most prior medical NLP used negation detection algorithms to deal with the negative cases. However, we found many negative cases that did not contain clear negative expressions. Therefore, we classified phenotype expressions as positive, negative, or neither (not relevant), and trained 3-class classifiers.

A few prior studies used NLP for delirium research. One such study [20] summarized patterns in the delirium literature over time, using unsupervised learning methods; by contrast, our work used NLP to extract information from clinical notes. Another study [21] detected delirium using an open-source NLP pipeline MedTaggerIE—an unstructured information management architecture–based information extraction framework. Shao et al [22] experimented with 3 different topic modeling methods and a keyword search method for identifying delirium-related documents and sentences in clinical notes. Weir et al [23] designed classifiers for patients with delirium by combining text data with ICD, Ninth Revision codes. Sun et al [24] defined a generic process for developing a clinical risk prediction model, applied the model calibration process at 4 hospitals, and generated risk prediction models for delirium. Jauk et al [25] implemented a random forest–based algorithm to identify hospitalized patients at high risk for delirium. A key difference between these prior studies and this study is that they aimed to detect delirium at the patient level (ie, whether a patient ever experienced delirium during a hospitalization). By contrast, our approach detects delirium at the sentence level, which provides more fine-grained temporal information (ie, on which days was a patient experiencing delirium). Such information is important for estimating the overall burden of delirium, and for studies that attempt to relate time-varying factors to the development of delirium.

Strengths

This work leveraged a large cohort composed of multiple different cohorts. These data sets provide a good source for variety of delirium expression in clinical notes. Additionally, we developed a novel GUI labeling tool and used active learning to enhance labeling efficiency. Furthermore, we compared 3 widely used NLP classifiers including a state-of-the-art Transformer model for delirium detection. Finally, we compared our delirium NLP detector with other delirium indicators, and we were able to demonstrate that our NLP method is substantially better than traditional methods based on ICD codes.

Limitations

Although our data were obtained from 9 hospitals, all were in the same geographic region (Massachusetts). Thus, our cohort may not be representative of other US or non-US populations. One important future direction is to test our delirium NLP algorithm using data from other regions. Additionally, the coverage rate of the “always pattern” for the development data set was 97.6% (n=195,680) owing to active learning, but decreased to 78.2% (n=12,569) on an independent test set. Further rounds of active learning to enlarge the available training data will help further expand the generalizability of the NLP Transformer model to new data sets. Nevertheless, our fine-tuning experiments show that extending the model to new data sets may require only a relatively small amount of additional labeling effort.

Conclusions

In this work, we developed a new delirium NLP detection approach that identifies patients with delirium from unstructured clinical notes. In many cases, the delirium information was only recorded in clinical notes and was absent from ICD codes. We anticipate that this model will be useful for large-scale EHR-based research on delirium, especially detecting delirium at a fine-grained level such as the note and sentence levels. Additionally, the labeling process based on active learning developed for this study was very efficient, achieving a coverage rate of 97.6% (n=195,680) in the development data set after just 5 rounds of labeling. This labeling method can be used for other studies related to phenotype detection based on unstructured clinical notes.

Acknowledgments

SD was supported by the National Institutes of Health (NIH; P30AG062421). SM was supported by the James S McDonnell Foundation (K23MH115812). MBW and WG were supported by the Glenn Foundation for Medical Research and American Federation for Aging Research (Breakthroughs in Gerontology Grant), the American Academy of Sleep Medicine (AASM Foundation Strategic Research Award), and the NIH (R01NS102190, R01NS102574, R01NS107291, RF1AG064312, and R01AG062989).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Demographic features.
[DOCX File , 19 KB - formative_v6i6e33834_app1.docx]

Multimedia Appendix 2
Examples for delirium sentences and always patterns.
[DOCX File , 87 KB - formative_v6i6e33834_app2.docx]

Multimedia Appendix 3
Two query strategies.
[DOCX File , 15 KB - formative_v6i6e33834_app3.docx]

Multimedia Appendix 4
Other delirium indicators.
[DOCX File , 18 KB - formative_v6i6e33834_app4.docx]

Multimedia Appendix 5
Interrater agreement.
[DOCX File , 13 KB - formative_v6i6e33834_app5.docx]

Multimedia Appendix 6
Data splitting.
[DOCX File , 24 KB - formative_v6i6e33834_app6.docx]

References

1. Fong TG, Tulebaev SR, Inouye SK. Delirium in elderly adults: diagnosis, prevention and treatment. Nat Rev Neurol 2009 Apr;5(4):210-220 [FREE Full text] [doi: 10.1038/nrneurol.2009.24] [Medline: 19347026]
2. Bucht G, Gustafson Y, Sandberg O. Epidemiology of delirium. Dement Geriatr Cogn Disord 1999 Aug 26;10(5):315-318. [doi: 10.1159/000017161] [Medline: 10473930]
3. Sepulveda E, Franco JG, Trzepacz PT, Gaviria AM, Meagher DJ, Palma J, et al. Delirium diagnosis defined by cluster analysis of symptoms versus diagnosis by DSM and ICD criteria: diagnostic accuracy study. BMC Psychiatry 2016 May 26;16:167 [FREE Full text] [doi: 10.1186/s12888-016-0878-6] [Medline: 27229307]
4. Kim DH, Lee J, Kim CA, Huybrechts KF, Bateman BT, Patorno E, et al. Evaluation of algorithms to identify delirium in administrative claims and drug utilization database. Pharmacoepidemiol Drug Saf 2017 Aug;26(8):945-953 [FREE Full text] [doi: 10.1002/pds.4226] [Medline: 28485014]
5. Spasic I, Nenadic G. Clinical Text Data in Machine Learning: Systematic Review. JMIR Med Inform 2020 Mar 31;8(3):e17984 [FREE Full text] [doi: 10.2196/17984] [Medline: 32229465]
6. McInnes L, Healy J, Saul N, Großberger L. UMAP: Uniform Manifold Approximation and Projection. JOSS 2018 Sep;3(29):861. [doi: 10.21105/joss.00861]
7. Tong S, Koller D. Support vector machine active learning with applications to text classification. J Mach Learn Res 2001;2:45-66 [FREE Full text] [doi: 10.1201/9781315371658-17]
8. Graves A. Supervised Sequence Labelling. In: Supervised Sequence Labelling with Recurrent Neural Networks. Berlin, Heidelberg: Springer; 2012:5-13.
9. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. Adv Neural Inf Process Syst 2007 Oct 01;45(02):45-0602-45-0602. [doi: 10.5860/choice.45-0602]

10. Devlin J, Chang M, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv. Preprint posted online October 11, 2018. [doi: 10.48550/arXiv.1810.04805]
11. Neamatullah I, Douglass MM, Lehman LH, Reisner A, Villarroel M, Long WJ, et al. Automated de-identification of free-text medical records. BMC Med Inform Decis Mak 2008 Jul 24;8:32 [FREE Full text] [doi: 10.1186/1472-6947-8-32] [Medline: 18652655]
12. Norgeot B, Muenzen K, Peterson TA, Fan X, Glicksberg BS, Schenk G, et al. Protected Health Information filter (Philter): accurately and securely de-identifying free-text clinical notes. NPJ Digit Med 2020;3:57 [FREE Full text] [doi: 10.1038/s41746-020-0258-y] [Medline: 32337372]
13. Davison A, Hinkley D. Bootstrap Methods and Their Application. Cambridge: Cambridge University Press; 1997.
14. Rogati M, Yang Y. High-performing feature selection for text classification. 2002 Presented at: CIKM02: Eleventh ACM International Conference on Information and Knowledge Management; November 4-9, 2002; McLean, VA p. 659-661. [doi: 10.1145/584792.584911]
15. McCoy TH, Castro VM, Roberson AM, Snapper LA, Perlis RH. Improving Prediction of Suicide and Accidental Death After Discharge From General Hospitals With Natural Language Processing. JAMA Psychiatry 2016 Oct 01;73(10):1064-1071. [doi: 10.1001/jamapsychiatry.2016.2172] [Medline: 27626235]
16. Gundlapalli A, South B, Phansalkar S, Kinney AY, Shen S, Delisle S, et al. Application of Natural Language Processing to VA Electronic Health Records to Identify Phenotypic Characteristics for Clinical and Research Purposes. Summit Transl Bioinform 2008 Mar 01;2008:36-40 [FREE Full text] [Medline: 21347124]
17. Zhou L, Baughman AW, Lei VJ, Lai KH, Navathe AS, Chang F, et al. Identifying Patients with Depression Using Free-text Clinical Documents. Stud Health Technol Inform 2015;216:629-633. [Medline: 26262127]
18. Yang X, Bian J, Hogan W, Wu Y. Clinical concept extraction using transformers. J Am Med Inform Assoc 2020 Dec 09;27(12):1935-1942 [FREE Full text] [doi: 10.1093/jamia/ocaa189] [Medline: 33120431]
19. Mascio A, Kraljevic Z, Bean D, Dobson R, Stewart R, Bendayan R, et al. Comparative analysis of text classification approaches in electronic health records. arXiv. Preprint posted online May 8, 2020. [doi: 10.18653/v1/2020.bionlp-1.9]
20. McCoy TH. Mapping the Delirium Literature Through Probabilistic Topic Modeling and Network Analysis: A Computational Scoping Review. Psychosomatics 2019;60(2):105-120. [doi: 10.1016/j.psym.2018.12.003] [Medline: 30686485]
21. Fu S, Lopes G, Pagali S, Thorsteinsdottir B, LeBrasseur NK, Wen A, et al. Ascertainment of Delirium Status Using Natural Language Processing From Electronic Health Records. J Gerontol A Biol Sci Med Sci 2022 Mar 03;77(3):524-530 [FREE Full text] [doi: 10.1093/gerona/glaa275] [Medline: 35239951]
22. Shao Y, Weir C, Zeng-Treitler Q, Estrada N. Identifying Documentation of Delirium in Clinical Notes through Topic Modeling. 2015 Presented at: 2015 International Conference on Healthcare Informatics; October 21-23, 2015; Dallas, TX. [doi: 10.1109/ichi.2015.47]
23. Weir C, LaFluer J, Gibson B, Zeng Q. Big Data Analytics Using the VA's 'VINCI' Database to Look at Delirium. In: Big Data-Enabled Nursing. Cham: Springer; 2017:287-299.
24. Sun H, Depraetere K, Meesseman L, De Roo J, Vanbiervliet M, De Baerdemaeker J, et al. A scalable approach for developing clinical risk prediction applications in different hospitals. J Biomed Inform 2021 Jun;118:103783 [FREE Full text] [doi: 10.1016/j.jbi.2021.103783] [Medline: 33887456]
25. Jauk S, Kramer D, Großauer B, Rienmüller S, Avian A, Berghold A, et al. Risk prediction of delirium in hospitalized patients using machine learning: An implementation and prospective evaluation study. J Am Med Inform Assoc 2020 Jul 01;27(9):1383-1392 [FREE Full text] [doi: 10.1093/jamia/ocaa113] [Medline: 32968811]

Abbreviations

AUC: area under the curve
BERT: Bidirectional Encoder Representations from Transformers
CAM: Confusion Assessment Method
EHR: electronic health record
GUI: graphical user interface
ICD: International Classification of Diseases
IRA: interrater agreement
LSTM: long short-term model
NIH: National Institutes of Health
NLP: natural language processing
RNN-LSTM: recurrent neural networks with LSTM units
ROC: receiver operating characteristic
SVM: Support Vector Machine

Edited by M Focsa; submitted 27.09.21; peer-reviewed by B Zanotto, Y Chen, Z Ren; comments to author 29.11.21; revised version received 22.01.22; accepted 10.02.22; published 24.06.22.

Please cite as:
Ge W, Alabsi H, Jain A, Ye E, Sun H, Fernandes M, Magdamo C, Tesh RA, Collens SI, Newhouse A, MVR Moura L, Zafar S, Hsu J, Akeju O, Robbins GK, Mukerji SS, Das S, Westover MB
Identifying Patients With Delirium Based on Unstructured Clinical Notes: Observational Study
JMIR Form Res 2022;6(6):e33834
URL: https://formative.jmir.org/2022/6/e33834
doi:10.2196/33834
PMID:35749214

Original Paper

Quantification of Digital Body Maps for Pain: Development and Application of an Algorithm for Generating Pain Frequency Maps

Abhishek Dixit[1], BSc (Hons), MRes; Michael Lee[1], MBBS, FRCA, PhD, FFPMRCA

Division of Anaesthesia, Department of Medicine, University of Cambridge, Cambridge, United Kingdom

Corresponding Author:
Abhishek Dixit, BSc (Hons), MRes
Division of Anaesthesia
Department of Medicine
University of Cambridge
Box 93, Addenbrooke's Hospital, Hills Road
Cambridge, CB2 0QQ
United Kingdom
Phone: 44 1223217 ext 889
Email: ad825@cam.ac.uk

Abstract

Background: Pain is an unpleasant sensation that signals potential or actual bodily injury. The locations of bodily pain can be communicated and recorded by freehand drawing on 2D or 3D (manikin) surface maps. Freehand pain drawings are often part of validated pain questionnaires (eg, the Brief Pain Inventory) and use 2D templates with undemarcated body outlines. The simultaneous analysis of drawings allows the generation of pain frequency maps that are clinically useful for identifying areas of common pain in a disease. The grid-based approach (dividing a template into cells) allows easy generation of pain frequency maps, but the grid's granularity influences data capture accuracy and end-user usability. The grid-free templates circumvent the problem related to grid creation and selection and provide an unbiased basis for drawings that most resemble paper drawings. However, the precise capture of drawn areas poses considerable challenges in producing pain frequency maps. While web-based applications and mobile-based apps for freehand digital drawings are widely available, tools for generating pain frequency maps from grid-free drawings are lacking.

Objective: We sought to provide an algorithm that can process any number of freehand drawings on any grid-free 2D body template to generate a pain frequency map. We envisage the use of the algorithm in clinical or research settings to facilitate fine-grain comparisons of human pain anatomy between disease diagnosis or disorders or as an outcome metric to guide monitoring or discovery of treatments.

Methods: We designed a web-based tool to capture freehand pain drawings using a grid-free 2D body template. Each drawing consisted of overlapping rectangles (Scalable Vector Graphics <*rect*> elements) created by scribbling in the same area of the body template. An algorithm was developed and implemented in Python to compute the overlap of rectangles and generate a pain frequency map. The utility of the algorithm was demonstrated on drawings obtained from 2 clinical data sets, one of which was a clinical drug trial (ISRCTN68734605). We also used simulated data sets of overlapping rectangles to evaluate the performance of the algorithm.

Results: The algorithm produced nonoverlapping rectangles representing unique locations on the body template. Each rectangle carries an overlap frequency that denotes the number of participants with pain at that location. When transformed into an HTML file, the output is feasibly rendered as a pain frequency map on web browsers. The layout (vertical-horizontal) of the output rectangles can be specified based on the dimensions of the body regions. The output can also be exported to a CSV file for further analysis.

Conclusions: Although further validation in much larger clinical data sets is required, the algorithm in its current form allows for the generation of pain frequency maps from any number of freehand drawings on any 2D body template.

(JMIR Form Res 2022;6(6):e36687) doi:10.2196/36687

KEYWORDS

Scalable Vector Graphics; SVG; pain drawing; pain location; Body Pain Map; overlap computation; heat map; pain frequency map; algorithm

Introduction

Background

Pain is an unpleasant sensation signaling potential or actual injury to the body site [1]. The location of pain can be communicated or recorded by drawing onto a body template (eg, Brief Pain Inventory [2]). Pain drawings have been used clinically and in research for decades [3]. Jang et al [4] reported that patients were more confident about communicating the locations of pain to the clinician in the form of a drawing as opposed to a written description, and the clinicians also favored drawings over written descriptions. Digital technology is now commonplace and circumvents problems associated with the processing and storage of pain drawings; hence, pain drawings are now widely acquired as digital images when possible [5]. Numerous web-based applications and mobile apps offer digital pain manikins, both commercial and academic [6]. Most digital pain manikins use 2D templates, which are whole-body coronal or sagittal views of the human body. There is often a choice of gender and body type for 2D or 3D surface templates [7-9]. A typical digital pain manikin consists of a body template and self-explanatory instructions for the patient regarding how to indicate the locations where they experience the most discomfort.

The pain frequency map is generated by the simultaneous analysis of all digital pain drawings to compute the locations of pain that participants have in common. To aid visualization, the maps also use color codes to highlight locations on the body in accordance with their frequency of occurrence in the sample being studied. Such maps are clinically useful in identifying where sensations commonly relate to disease anatomy [10,11] and the factors that influence the subjective localization of pathology. These maps are perhaps most important in chronic primary pain syndromes [12], which are defined by the locations of pain in the absence of disease (eg, chronic back pain).

The ease with which pain frequency maps are generated critically depends on the nature of the body template used. For body templates where the anatomical locations are already predefined and demarcated [7,13], the pain drawings are swiftly completed, and the data captured are binary, where *1* indicates a selection, and *0* indicates no selection. As a result, the generation of pain frequency maps is relatively straightforward, because only the number of participants who selected each location needs to be deduced [14]. However, such frequency maps may not entirely capture the underlined spatial patterns of pain because of loss of spatial resolution caused due to the pixel (ie, x-y coordinate, representing the smallest possible division for the body template) being preassigned to larger (hence fewer) anatomical regions (eg, the Collaborative Health Outcomes Information Registry [CHOIR] body template has 74 divisions to click [7]). In addition, assumptions must be made about where the anatomical locations of pain are clinically relevant or important.

Freehand pain drawings, using body templates with undemarcated or blank body outlines, are unbiased. Some body templates supporting freehand drawings use a grid with a predefined number of cells (eg, GeoPain app with a 3D body template uses a grid of 2026 cells [9] and Manchester Digital Pain Manikin uses a grid of 12,800 cells for its 2D body template [15]). The *grid-based* templates also allow the easy generation of pain frequency maps because each location (a cell) is binary in nature and is either selected or not selected. However, as body templates come in a variety of shapes and sizes [5], the aspect ratio (ratio of width to height) of the body regions also changes, and it is therefore not possible to standardize a grid (no *one-size-fits-all*). The optimal grid granularity (resolution) of each body template must be assessed. The granularity of the grid determines the end-user experience and accuracy of the captured data [15].

The *grid-free* body template, as the name suggests, overcomes the problem of grid creation and selection and provides an assumption-free basis for pain drawings that most resemble paper drawings. The number of clickable locations is of the order of thousands (depending on the pencil size). In addition, the participant may choose to draw repeatedly in the same location, similar to what happens when pen and paper are used. The precise capture of drawn areas, along with no predefined locations (cells), poses considerable challenges in generating pain frequency maps. The drawings on grid-free templates require intricate pixel-level analysis to generate a pain frequency map, and the tools for generating such maps are lacking or not freely available [16-18].

Objectives

In this paper, we describe a novel and unbiased algorithm developed specifically to generate a spatial pain frequency map from freehand pain drawings on a generic 2D whole-body template. We also assessed the performance of the algorithm's Python script and demonstrated its utility by generating pain frequency maps from pain drawings obtained from 2 clinical data sets.

Methods

Creating the Digital Body Template (Manikin)

An image outline (or template) of the human body is embedded in an HTML page to create a digital body template (manikin) to capture drawings of pain locations.

The responsiveness of the body template (ie, the ability to highlight or zoom in or out of a region) requires vectors, which are created using lines, points, and shapes to represent the different regions (demarcations) of the body. Using Scalable Vector Graphics (SVG), an XML-based language, it is possible to display vectors and create a responsive body template.

The locations of pain can be recorded on any body template by inserting basic SVG elements such as circles and rectangles. These shapes are inserted by specifying their position and size as the core attributes. For example, the circle requires coordinates (*cx*, *cy*) of the center and the radius (*r*), and the rectangle requires coordinates (*x*, *y*) of the top left corner along with *width* and *height*. In addition to the core attributes, the SVG elements can also contain style-related attributes (eg, fill, visibility, and opacity) and any number of customized attributes with prefix *data-* (eg, data-region and data-date-inserted).

The circle and rectangle elements can be created in SVG:

<circle cx="5" cy="5" r="10" fill="green" visibility="hidden" />

<rect x="5" y="5" width="20" height="10" fill="red" opacity="0.8" data-region="front" />

We used the SVG *<rect>* element to record the location of pain because the intersection of the 2 rectangles is always a rectangle (consistent geometry) that is required for the algorithm to function.

We downloaded a sexless human body outline image (TIFF), which is an adaptation of one of the oldest templates from the early works of Sir Henry Head [19]. The image has 4 views of the human body in 2D: coronal (front and back) and sagittal (left and right). The image was modified to demarcate 4 nonoverlapping regions: the front, side head (right), side head (left), and back. For making regions *clickable*, the body image was vectorized and converted into SVG using the Inkscape Editor (version 1.1) [20]. Vectorization produced 4 vectors (beziergons), where each vector has an associated bounding box, which is the tightest fitting rectangle that encloses all points on the vector (Figure 1).

In order to facilitate data capture from drawings of bodily pain locations, we first developed *Body Pain Map*, a tool using Linux-Apache-Perl-MongoDB–based infrastructure. The HTML5 webpage of the tool consists of a 2-column layout where the right column embeds the SVG body template and the left column contains 7 JavaScript-powered control elements (Figure 2), which are (1) size of rectangular pencil tip (*<select>* element with 3 options, small, medium, and large, signifying red squares of dimensions: 10, 30, and 60). The dimension of the smallest square was based on the smallest width found on the body template (eg, little finger), (2) zoom in (*<button>* element) to allow a closer view, (3) zoom out (*<button>* element) to allow a wider view, (4) Erase (*<button>* element) to remove a previous pain recording, (5) undo eraser (*<button>* element) to restate a previously removed recording, (6) clear (*<button>* element) to remove all recordings, and (7) submit (*<button>* element) to submit drawings for storage in the MongoDB database for analysis and reconstruction purposes. A blank body template can be submitted (eg, when the participant has no bodily pain to draw).

For making the Body Pain Map easily accessible as a web-based tool, we placed the HTML document on a secure Linux machine running the Secure Socket Layer (SSL)–enabled Apache HTTP web server (Figure 3). The tool can be accessed on the web [21] and tested on modern web browsers (eg, Google Chrome, Mozilla Firefox, and Microsoft Edge).

Figure 1. The body template (manikin) shows 4 regions. The tightest fitted rectangle (bounding box) enclosing each region is shown in black.

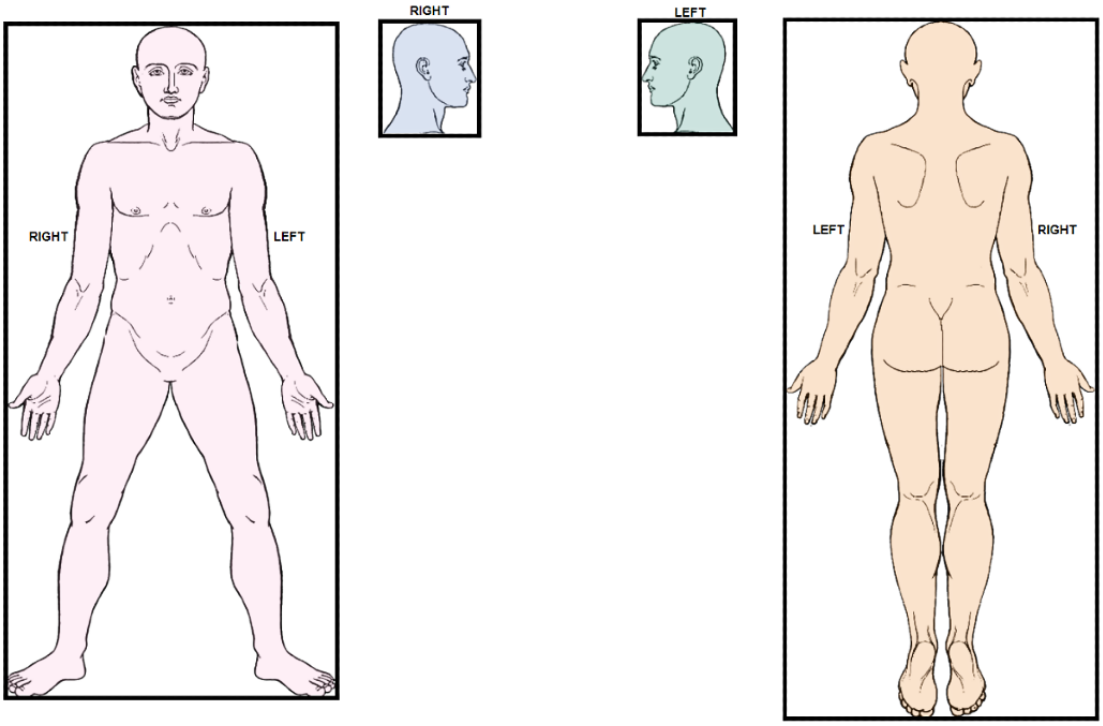


Figure 2. HTML webpage with a 2-column layout. The left column contains all control elements (buttons) and the right column contains the Scalable Vector Graphics body template. The participant can click anywhere on the template (divided into 4 regions) to locate their pain; a pain drawing of body regions, which comprises overlapping squares (rectangles of same width and height) of varying sizes is shown here. These drawings are input to the algorithm.

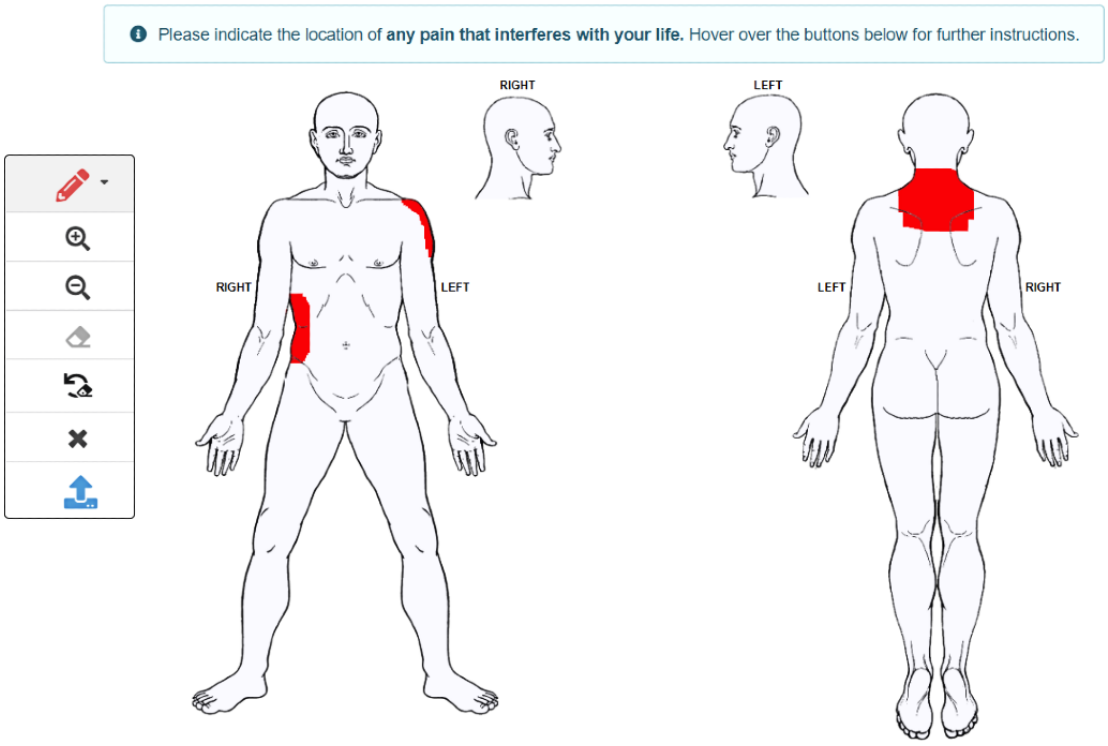
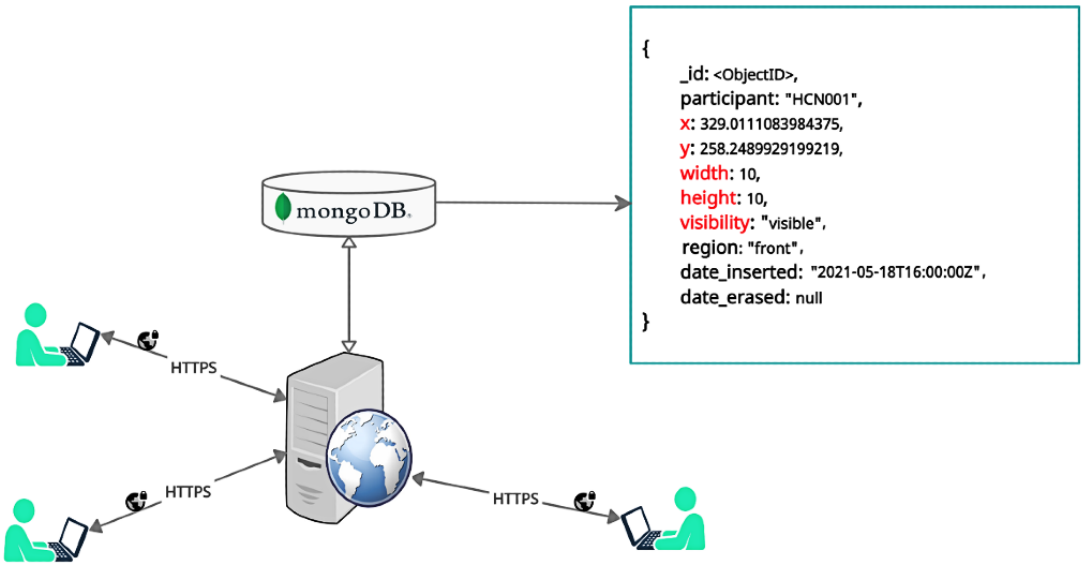


Figure 3. Each participant accesses the Body Pain Map tool through their web browser securely. The pain drawings are captured and stored in a MongoDB database. A sample JSON document of the pain recording (denoting Scalable Vector Graphics <rect> element) is shown as an example with core attributes in the color red.



Developing the Algorithm to Generate a Pain Frequency Map

Overview

The pain frequency map is generated by superimposing several drawings made by the participants using a rectangular pencil tip on a given body template. The drawing from each participant indicates where the pain is located on the body and can comprise multiple overlapping rectangles. The degree to which the rectangles overlap when all drawings are superimposed is of interest. The algorithm described in the *Overlap Computation Algorithm* section generates a pain frequency map that denotes the proportion of overlap in the rectangles between participant drawings. By default, the areas where the proportion of overlap is higher (ie, where pain is more commonly located) are redder in color. Areas with less overlap appear less red. Areas that are colored white are regions within the body template where no participant has drawn.

Overlap Computation Algorithm

The key steps in the overlap computation algorithm are (1) data decomposition, (2) creating partitions, (3) merging partitions to create nonoverlapping rectangles, and (4) optimizing nonoverlapping rectangles.

Data Decomposition

The first step toward 2D overlap computation is the decomposition of the source data from all participants' drawings on a given 2D body template.

Let T be the total number of participants, P ($|P| \leq T$) be the set of strings denoting all participants with a drawing, R be the set of visible and unique rectangles from all participants in the *front* region of the body template, and I be the index set of set R, we assume that all participants in P contribute a drawing each, and as each participant drawing contains at least one rectangle, we define the relationship between rectangles (indexes) and participants as a surjective function *rectToParticipant*:

Assuming that each rectangle has attributes x (x coordinate of the top left corner), y (y coordinate of the top left corner), w (width) and h (height) we decompose rectangles along the x- and y-axis using the functions *coordX* and *coordY*:

Creating Partitions

In this step, we create nonoverlapping partitions along the x- and y-axis. Each partition is an interval window of varying size and represents an area enclosed by at least one rectangle.

Let E be the domain and C (family of sets) be the range of the function *coordX* or *coordY* depending on the axis.

The function *partition* for creating partitions along the x- and y-axis is defined as follows:

Merging Partitions to Create Nonoverlapping Rectangles

This step involves merging the X and Y partitions (nonoverlapping interval windows) to create nonoverlapping rectangles.

Let Px and Dx (family of sets) be the domain and range of *partitionX* and Py and Dy (family of sets) be the domain and range of *partition.*

We define O (a family of sets of participants P) as a set of all overlaps. The function *rectToParticipant* defined earlier is used to map the rectangles to their corresponding participants.

For any o ($o \in O$) the set of nonoverlapping rectangles R_{ME} can then be defined as

Each element in R_{ME} represents a unique, nonoverlapping location in the *front* region of the body template and carries an overlap frequency $|o|$ or proportion given by  .

Optimizing Nonoverlapping Rectangles

In the previous step we generated nonoverlapping rectangles some of which might be adjoined.

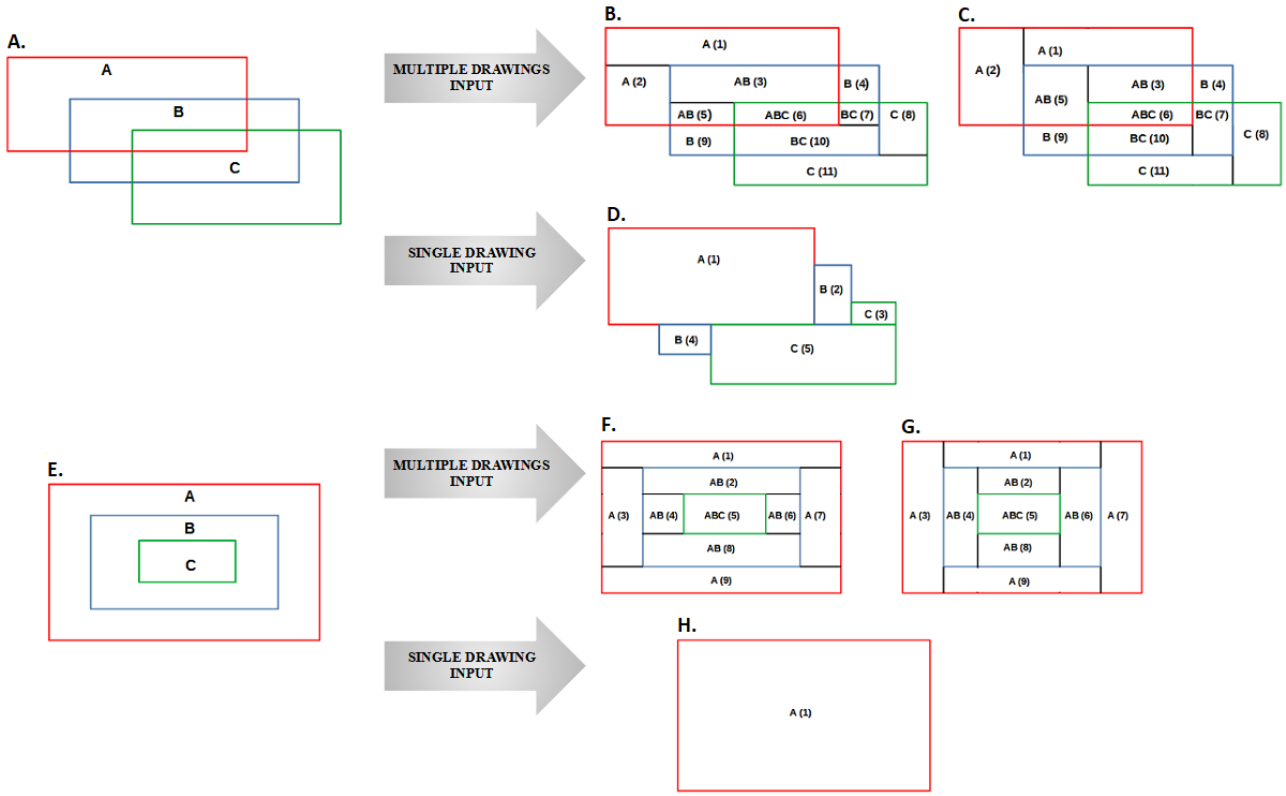
This step describes the merging of adjoining rectangles for all the observed overlaps (O). The merging of rectangles is important for two main reasons: (1) to optimize the dimensions of the rectangles (ie, as wide or long as possible) and (2) to reduce the size of the output file (SVG) for efficient rendering by the web browser.

Let r_1 and r_2 be 2 nonoverlapping rectangles associated to the overlap o ($o \in O$).

The 2 rectangles r_1 and r_2 are eligible for horizontal or vertical merging, provided that either of the following conditions is met.

Horizontal merging produced wider rectangles, whereas vertical merging produced longer rectangles (Figure 4).

Figure 4. Illustration of the algorithm's input and output. (A) A, B, and C are individuals who have drawn red, blue and green rectangles that overlap on a body template. (B) These overlapping rectangles are inputted to the algorithm. The output is 11 nonoverlapping rectangles (horizontal layout), named 1 to 11 in parentheses. Each rectangle represents a location with the proportion of pain. For example, A(1) = 1/3 , AB (3) = 2/3, ABC (6) = 3/3. (C) This shows the output from the algorithm in the vertical layout. The choice of layout (horizontal or wider rectangles or vertical or longer rectangles) can be made as per the locations of pain. For the legs, the rectangles might be better visualized when the layout is vertical, whereas horizontal layout is preferred for the abdominal area. (D) The illustration of the algorithm's output when the input is a single drawing made by one individual (A, B and C are the same individual). In this case, the output is simply 5 nonoverlapping rectangles instead of 11 each with a proportion of 1.0. (E) A, B and C are individuals who have drawn red, blue and green rectangles that overlap in a nested fashion, on a body template. These nested rectangles when inputted to the algorithm produce 9 nonoverlapping rectangles, with (F) showing the horizontal and (G) showing the vertical layout. (H) The demonstration of the output (ie, 1 nonoverlapping rectangle) when the input is a single drawing consisting of nested rectangles. This tends to occur when the individual elects to switch between pencil (tip) size: small, medium and large.



Algorithm Analysis Pipeline Construction

For implementing the overlap computation algorithm, we developed a command line workflow in Python programming language release 3.9 [22], which involves 3 steps (scripts) described further. The scripts can be run in either sequential or pipeline mode (using the named pipe command "|"). Each script performs a specific operation and accepts parameters in the form of command line options and arguments. The scripts use some core Python modules that are responsible for reading command line arguments, parsing and validating SVG to be processed, and producing the output for the subsequent step in the workflow. The following scripts can be obtained by contacting the corresponding author (AD):

1. *extract_data.py*: this script is only used if the pain drawings are originally saved as SVG files (a file per participant). The script extracts all the pain recordings denoted by *<rect>* elements with the same width and height attributes. The script produces a CSV output. The columns in the output are (1) *participant*: the identifier of the drawing, (2) *x*: the x coordinate of the top left corner of the rectangle, (3) *y*: the y coordinate of the top left corner of the rectangle, (3) *width*: the width of the rectangle, (4) *height*: the height of the rectangle (same as width), (5) *region*: the region of the body template where the recording belongs (eg, front or back), and (6) *visibility*: the visibility status of the recording (hidden means erased). The script also provides the option of including or excluding *empty* files. These files represent instances when the participant had no pain to draw or indicate on the body template. The script also allows users to extract data from specific body regions (eg, front or back).
2. *compute_overlap.py*: this script implements the overlap computation algorithm to produce nonoverlapping rectangles. Each rectangle represents a unique location on the body template. The script accepts input in CSV format with columns, namely, *participant*, *x*, *y*, *width*, *height*, and *region*. The script returns a CSV output where rows are nonoverlapping rectangles, and the columns (rectangles' attributes) are (1) *x*: the x coordinate of the top left corner of the rectangle, (2) *y*: the y coordinate of the top left corner of the rectangle, (3) *width*: the width of the rectangle, (4) *height*: the height of the rectangle, (5) *area*: the area of the rectangle, (6) *overlap*: the identifiers of drawings that overlap, (7) *overlap_frequency*: the number of identifiers that overlap, and (8) *overlap_proportion*: the proportion of identifiers that overlap. The proportion is calculated from the total number of drawings, which may include empty

drawings. The script also allows users to filter the output by providing the range for overlap (frequency) and thresholds for the width and height of the rectangles.

3. *plot_heatmap.py*: this script plots the nonoverlapping rectangles on the blank SVG body template and generates a pain frequency map (heat map) as an HTML file. The color of the pain frequency map can be specified either by its native name or as a hexadecimal color code (default #ff0000 or red). The intensity (shade or gradient) of the participant overlap on the pain frequency map is displayed using the opacity attribute of an SVG element. Opacity is any number strictly between 0 and 1. For optimum color coding, the opacity for each output rectangle is calculated as follows:

Data Generation for Algorithm Implementation

Simulation

We examined the performance of the algorithm's Python script *compute_overlap.py* in 2 separate simulations on a machine (Intel Xeon[R] Silver 4110 CPU@2.10GHz and 16-GB RAM) running Ubuntu 18.04. Given an XY plane of dimensions 1000×1000 with the origin at (0,0), let X and Y be the set of all natural numbers on the x- and y-axis. The set of all ordered pairs P is denoted by the Cartesian product $X \times Y$ of sets X, Y.

Simulation 1

We assumed that the drawings from every individual consisted of a single rectangle (analogous to a single mouse *click* on the body template). This situation is extremely unlikely; however, the purpose of this simulation was to test the ability of the algorithm to compute an overlap, given a set of overlapping rectangles.

A total of 10 data sets were generated for this simulation. The first data set consisted of 10,000 rectangles, and for each consecutive data set, the number of rectangles was increased by 10,000. For each data set, the coordinates of the origin of the rectangles were sampled without replacement from the set of ordered pairs (P), and the dimensions (width and height) were sampled with replacement from a sequence starting at 10 and ending at 100 (incremental step is 10).

Simulation 2

We assumed that a typical participant drawing of pain locations consisted of 100 rectangles (equivalent to 100 mouse *clicks* on the body template). For this simulation, we sought to assess performance with an increasing number of participants.

A total of 10 data sets were generated for this simulation. The first data set consisted of 100 participants, and for each consecutive data set, we increased the number of participants by 100. For each participant, the origin of the rectangles was sampled without replacement from the set of ordered pairs (P), and the dimensions (width and height) were sampled with replacement from a sequence starting at 10 and ending at 100 (incremental step is 10).

Each simulation produced 10 CSV files for analysis by the algorithm. The CSV file consisted of the columns, namely, *participant*, *x*, *y*, *width*, *height*, and *region*. For both simulations, there was only one region (the XY plane); therefore, the *region* was simply labeled *xy-plane*.

Actual

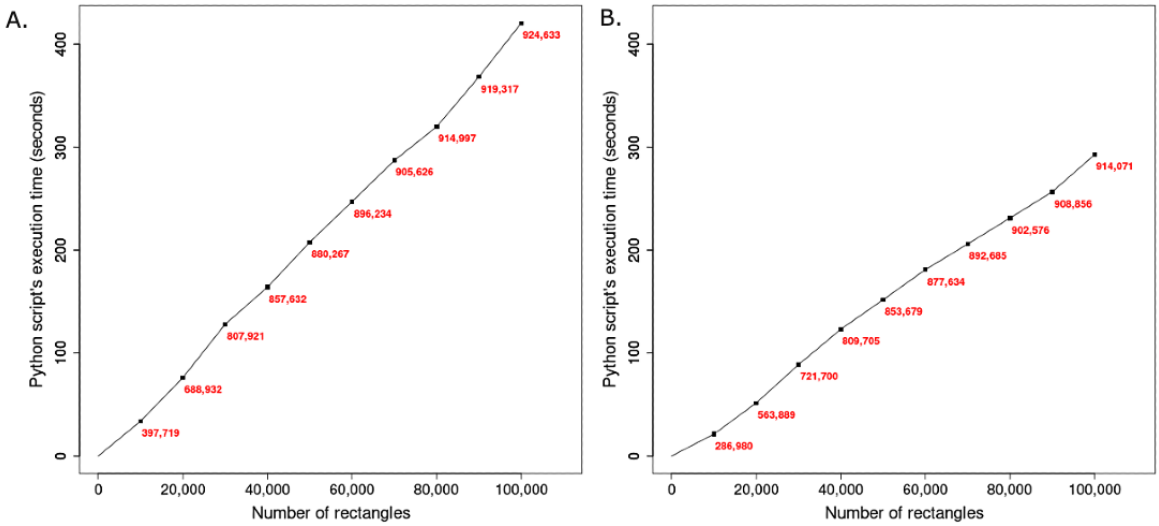
Finally, we used the digital pain drawings from 2 clinical data sets; the first data set (data set 1) consisted of 23 individuals who were screened for a clinical drug trial (ISRCTN68734605) [23], and the second data set (data set 2) comprised 30 participants with chronic back pain who were recruited into a separate study [24]. The pain drawings were stored as MongoDB documents (Figure 3), which were organized into subdocuments ordered by the date and time of insertion. The virtual pencil tip used for the drawings was a square, an SVG <*rect*> element, where the width and height attributes were the same (Figure 2).

Results

Simulation

The simulations were carried out to assess the performance of the algorithm's Python script *compute_overlap.py*. The simulations reveal that the script's execution time (seconds) increases linearly as a function of the number of input rectangles (Figure 5), up to 400 seconds for 100,000 rectangles (simulation 1). Assuming that a typical participant drawing consists of 100 rectangles, the execution time for 1000 participant drawings is 300 seconds (simulation 2).

Figure 5. Simulations to assess the performance of the algorithm's Python script – compute_overlap.py. The y-axis shows the execution time of the script in seconds and the x-axis shows the number of rectangles handled by the script. The output from the script, the number of nonoverlapping rectangles is shown in red. The plot (A) shows the script's performance on the data sets from simulation 1 where the number of rectangles per participant was set to 1, and plot (B) shows the performance on the data sets generated in simulation 2 where the number of rectangles per participant drawing was set to 100.



Actual

We first extracted all drawings (<rect> elements with the same width and height attributes) from the 2 data sets stored in the MongoDB database into CSV files. Data set 1 produced 4016 and data set 2 produced 4167 recordings, respectively, sorted by the participant identifier and the date and time of insertion. The characteristics of the participants from the 2 data sets pertaining to the *Body Pain Map* exercise are summarized in Table 1.

For generating pain frequency maps, we only used visible and unique (nonidentical) recordings from the 2 clinical data sets. Identical recordings from the same participant (<*rect*> elements with complete overlay) were filtered based on the recording's last visibility status. If visibility was hidden (meaning erased), the recording was excluded. This exercise produced 2 CSV files (one file per data set) with 3242 and 3993 rows, respectively, signifying unique and visible recordings across all 4 regions, namely, front, side (right), side (left), and back as summarized in Table 2.

The CSV data sets were independently processed by the Python script *compute_overlap.py* on a machine (Intel(R) Xeon (R) Silver 4110 CPU @ 2.10GHz and 16 GB RAM) running Ubuntu 18.04. The script processed the first CSV data set in 4 seconds and produced 6653 nonoverlapping rectangles with an overlap frequency between 1 and 8. The second CSV data set was processed in 7 seconds and produced 6010 nonoverlapping rectangles with an overlap frequency ranging from 1 to 21. Each nonoverlapping rectangle represents a unique location on the 2D body template and carries an overlap frequency, which denotes the number of participants with pain at that location.

The nonoverlapping rectangles for the 2 clinical data sets were subsequently plotted on the body template using the Python script *plot_heatmap.py*, and a pain frequency map was generated for data set 1 (Figure 6) and data set 2 (Figure 7).

Table 1. Pencil and eraser clicks made by the participants from the 2 clinical data sets while using the Body Pain Map tool.

Characteristic	Data set 1	Data set 2
Participants, N	23	30
Clicks, mean (SD)		
Pencil[a]	174.6 (208.5)	138.9 (108.2)
Eraser[b]	16.7 (32.7)	1.5 (4.2)
Visible and unique pencil clicks, mean (SD)	141.0 (176.5)	133.1 (105.5)
Mean pencil size[c], mean (SD)	26.6 (10.3)	25.2 (6.0)

[a]Each pencil click denotes a recording of the pain location and creates a Scalable Vector Graphics <*rect*> element with the same width and height attributes (a square).

[b]The eraser click hides the previously inserted Scalable Vector Graphics <*rect*> element.

[c]On the basis of the visible and unique pencil clicks to record (draw) pain locations.

Table 2. Visible and unique pencil recordings made in each region for 2 clinical data sets[a].

Characteristic	Data set 1	Data set 2
Recordings, N	3242	3993
Region, n (%)		
Front	1381 (42.6)	800 (20.0)
Side (right)	80 (2.5)	0 (0.0)
Side (left)	103 (3.2)	11 (0.3)
Back	1678 (51.7)	3182 (79.7)

[a]Data are derived from the Scalable Vector Graphics drawings and provided to the algorithm's Python scripts to generate a pain frequency map.

Figure 6. Illustration of the pain frequency map produced from freehand pain drawings (data set 1) obtained from patients (N=23) who were screened for a clinical drug trial. In the interactive map (HTML format), the user can slide the black pointer on the gradient bar (shown in the middle) to view locations based on the overlap threshold (ie, <= overlap proportion) and save the map as a PNG file.

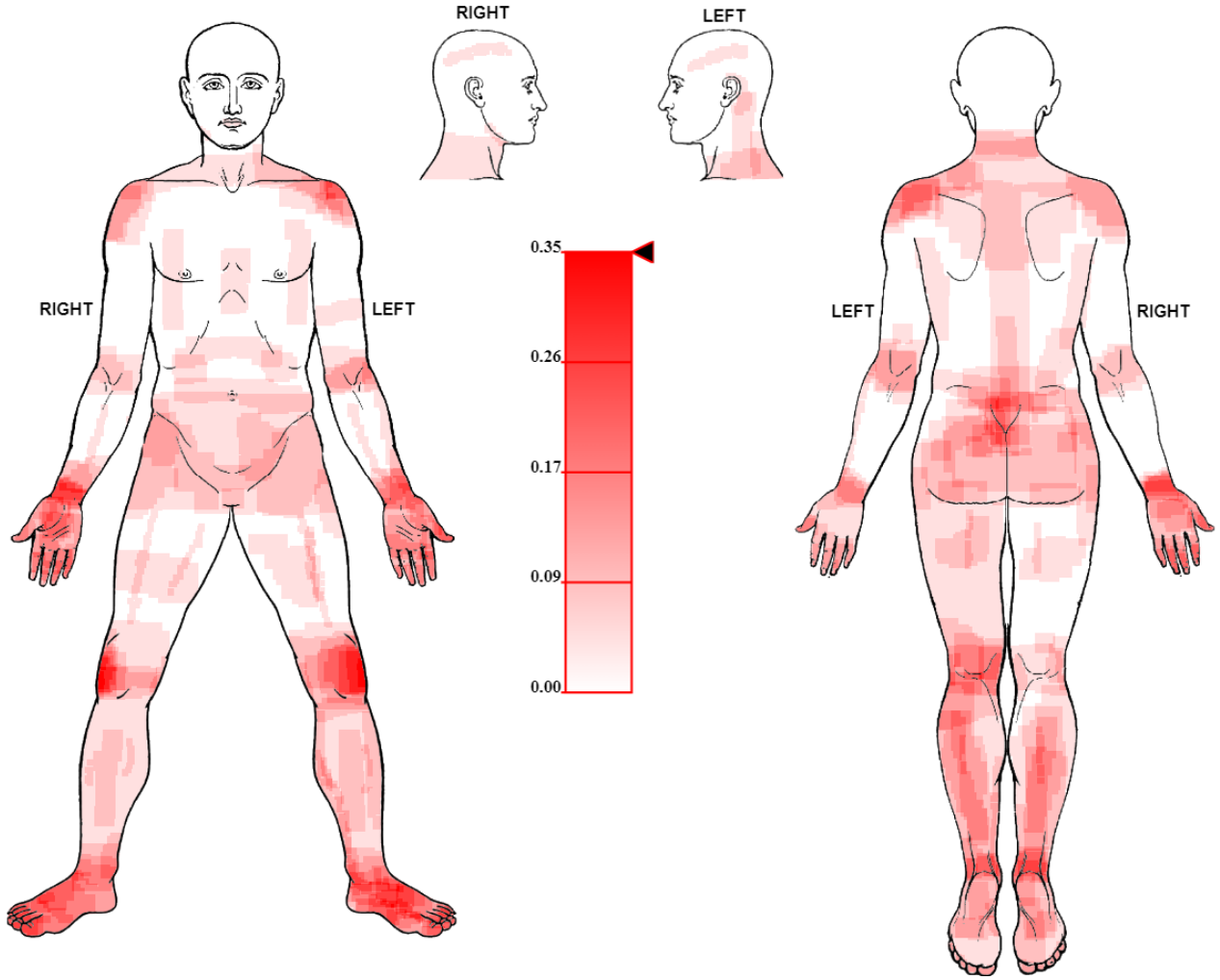
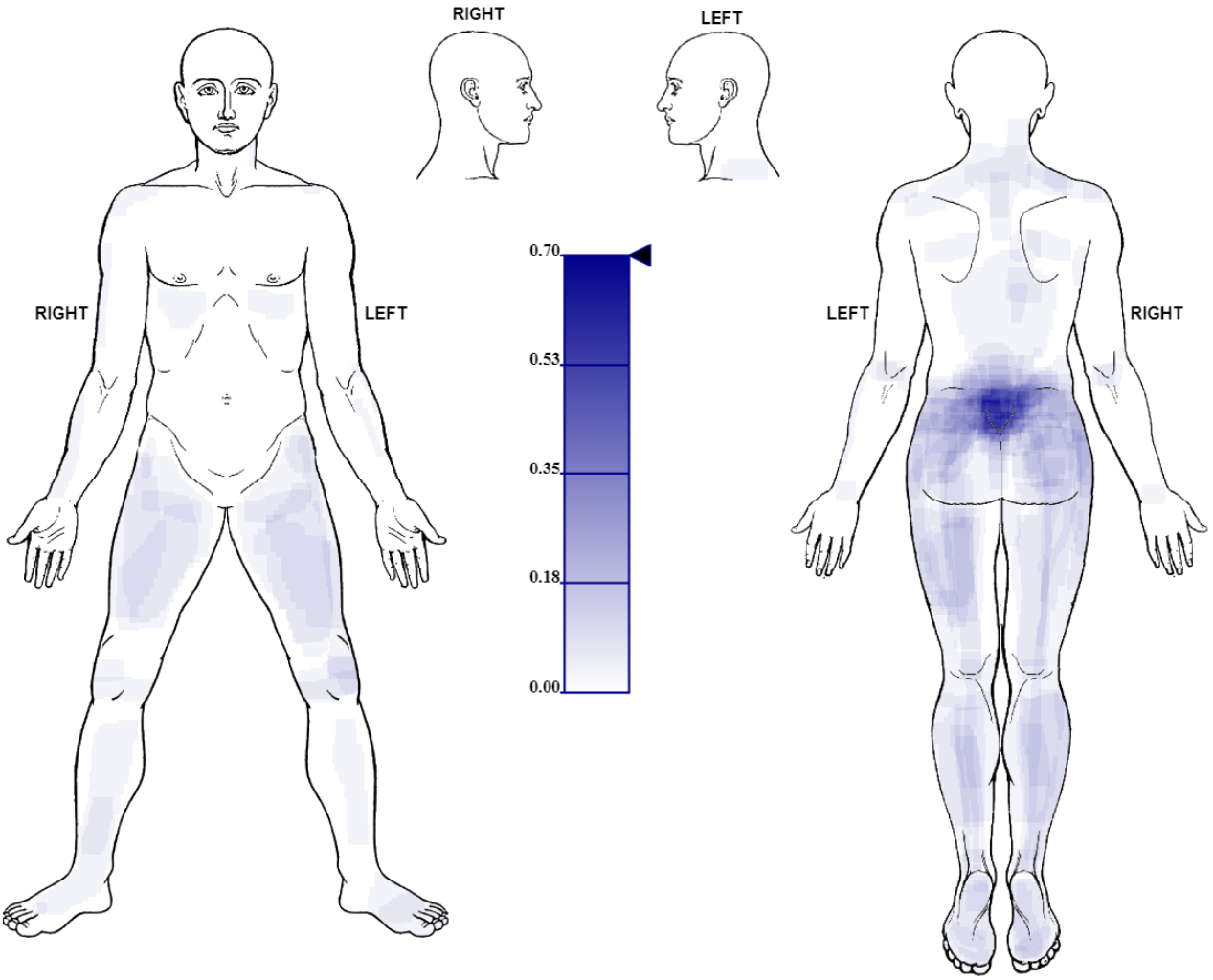


Figure 7. Illustration of the pain frequency map produced from freehand pain drawings (data set 2) obtained from patients (N=30) diagnosed with chronic primary back pain.



Discussion

Principal Findings

Drawing pain locations on body templates is widely and increasingly being used in research and pain clinics. These are frequently part of questionnaires (eg, the Brief Pain Inventory [25]). The topography of bodily pain is often summarized as a pain frequency map but relies on digitization of paper drawings, which is labor intensive and may be infeasible for larger studies [26]. Although freehand digital drawing tools are available to capture pain locations, they are often restricted to specific body templates [6]. Other tools parcellate the body (eg, the CHOIR Body Map [7] and Michigan Body Map [13]), which allows easy generation of pain frequency maps but lacks the resolution required.

In this paper, we first described the Body Pain Map [21], a web-based platform used to capture freehand pain drawings as SVG *<rect>* elements and then thoroughly described the algorithm to generate a pain frequency map from the drawings. We further described the implementation of the algorithm as a Python-based analysis pipeline and tested its performance in 2 different simulations. We demonstrated the utility of the algorithm by producing pain frequency maps (Figure 6 and Figure 7) from freehand drawings obtained from 23 patients screened for a clinical drug trial (ISRCTN68734605) [23] and 30 patients diagnosed with chronic primary back pain [24].

The key advantage of this algorithm is that it can handle data from any undemarcated body template. In other words, the body template does not need to be divided using a mesh with a fixed number of cells (eg, GeoPain, a body surface map rendered on a 3D manikin [9] and Manchester Digital Pain Manikin [15]) or demarcated into regions with anatomical labels (eg, CHOIR [7]). Only the outline or perimeter of the body template requires demarcation. The label-free and grid-free approach is unbiased and allows the participants to freely draw the locations of their bodily pain, closely mimicking the experience of drawing on paper. Although the user is unlikely to draw in the entire space, given the nature of freehand drawing, they can draw repeatedly in the same area. This results in redundant data, which are not usually of interest to most clinician researchers. Our algorithm processes raw (source pixel) data in grid-free space to address this difficulty, and in doing so, it also achieves lossless compression.

The algorithm accepts a single CSV (plain text) file with pain recordings captured as rectangles and produces nonoverlapping rectangles. Each nonoverlapping rectangle is a unique location

on the body template and carries an overlap frequency, which denotes the number of participants with pain in that location. The output, when transformed into an HTML file, can be feasibly rendered on modern web browsers (eg, Google Chrome, Mozilla Firefox, or Microsoft Edge) or printed at the desired resolution for publishing as we have shown for 2 clinical data sets. The algorithm also contains features to optimize the display of the pain frequency map. Rectangles can be produced either in a horizontal (wider) or vertical layout (longer; Figure 4). For example, for the leg, the rectangles may be better visualized with a vertical layout, whereas a horizontal layout is preferred for wider regions such as the abdomen.

As the individual can draw freely on the body template, the output (ie, nonoverlapping rectangles) produced by the algorithm is highly granular. Given the required spatial information (x-y coordinates within the body region), it is possible to retrospectively create a grid and reassign the cells and anatomical labels as required. The output from the algorithm can be subjected to coordinate transformation using linear algebra and labeled to allow harmonization with other body atlases or templates with anatomical labels (eg, arms and legs) in nature [7]. The unbiased output can also be trained to provide demarcations for a given pain disorder within any 2D body template. The demarcations or boundaries of areas for an anatomical label (eg, shoulder) may also be determined empirically through feedback from any group of individuals. It is also possible to acquire drawings from gender-specific body templates provided by the Hannover Medical School [8] and subsequently use the coordinate transformation to harmonize the algorithm's output on a gender-neutral template for comparisons between sexes.

Limitations

The algorithm generates nonoverlapping rectangles without any knowledge of the region boundaries within the 2D body template. Hence, some rectangles in the output may fall partly outside these regions. This overflow occurs when drawings include the border (or are close to the border) of the body regions. For generating pain frequency maps, this is not a problem because rectangles falling outside the regions can be easily masked (hidden) using SVG. For other analysis purposes, the problem can be feasibly mitigated by approaches such as (1) dividing the body template into small regions and applying bounding box correction to filter output nonoverlapping rectangles by region or (2) creating a *null* drawing that delicately fills the entire body template, ensuring that the <*rect*> elements are within the regional boundaries. With this approach, the *null* drawing is added to actual participant drawings, and the algorithm is executed on the combined data set. In the resulting output, only the nonoverlapping rectangles that are common to both the *null* drawing and the actual drawings are retained, and the rest (overlap frequency is 1) are discarded.

The algorithm is optimized to generate pain frequency data only for the locations of the body template on which at least one participant has drawn. Adding the *null* drawing would allow representation (in the pain frequency map) of body locations within the template that no participant has drawn and may also be required for inferential statistics. As the *null* drawing is specific to a body template, it does not have to be recreated, and the approach can be applied to prospective data sets (drawings), provided they are acquired using the same body template.

Comparison With Prior Work

Software tools (R package, CHOIR Body Map [27]) have recently been developed to generate co-occurrence maps. The map shows the number of times 2 locations on the body template are endorsed together by patients in a data set. However, these tools are only applicable to the CHOIR body template [7], where the participant can only click on 74 predefined locations.

Generating a pain frequency map for body templates with either demarcations or grid is straightforward because the locations are predefined and fixed, and it is a case of deducing the number of participants that selected each location [14].

Studies in which a nondemarcated and grid-free body template was used for freehand drawings used customized tools for the simultaneous analysis of their drawings. These tools are not freely available in the public domain [16-18]. Other studies [28,29] converted their drawings (originally saved as PNG images) into NIfTI format [30] and analyzed them using image-processing tools provided by the Functional Magnetic Resonance Imaging of the Brain (FMRIB) Software Library [31].

The primary purpose of our algorithm and other methods used in previous studies is to generate pain frequency maps, which requires simultaneous analysis of all freehand drawings provided by participants. Previous studies have stored freehand pain drawings as bitmap images (eg, PNG); therefore, the generation of a pain frequency map requires the extraction and analysis of all pixels [16-18,28,29].

Our algorithm processes drawings in which pain locations are indicated using rectangles (the pencil tip is a square, which is fundamentally a rectangle). This has several advantages, such as the fact that the input from the drawings can be simply stored as a CSV (plain text) file because all locations are represented as rectangles with attributes *x*, *y*, *width*, and *height*. This is also more efficient than the storage and extraction of pixels from the drawings. The output also comprises rectangles and can be stored as CSV files for statistical analyses and visualized using images (eg, SVG and PNG). We used SVG because it allows the reconstruction and visualization of the input and output rectangles at the desired resolution. CSV storage also facilitates the merging of several independent data sets acquired using the same body template for a combined analysis.

Conclusions

Body maps have long been used in research and clinical practice to facilitate communication between pain and other sensations [32]. The choice of a drawing tool or body part selection depends on the nature of the research question and the participants. It is important to clinically validate any digital tool used to capture the topography of body sensations [33]. Our algorithm is primarily developed to render pain frequency maps for efficient display and printing, but the output (ie, nonoverlapping rectangles) can be readily subjected to statistical

analyses (eg, statistical comparisons of pain frequency maps between different patient cohorts or the same patient cohort over multiple time points). Although we chose to digitize and use a specific body template in our study, the algorithm described can process any number of freehand drawings on any 2D body template to produce a pain frequency map. The nonoverlapping rectangles generated by the algorithm can be labeled anatomically or mapped onto a grid to facilitate analyses and harmonization with other body templates.

Our freehand pain drawing tool (Body Pain Map [21]) uses resolution-independent and XML document object model–based SVG technology. However, our algorithm can also generate pain frequency maps from drawings created using other technologies (eg, HTML5 <*canvas*> element), provided that the pain locations are captured as rectangles and the location attributes (ie, *x*, *y*, *width*, and *height*) are accessible.

As the algorithm has already been implemented as a Python command line workflow, it is possible to schedule an automated pain frequency map construction through the cron daemon (Linux environment) and filter and visualize the output using several criteria (eg, region, overlap frequency, width, and height of rectangles). The Python scripts can be obtained by contacting the corresponding author (AD).

We envisage the use of the algorithm in clinical or research settings to facilitate fine-grain comparisons of human pain anatomy between disease diagnosis or disorders or as an outcome metric to guide the monitoring or discovery of treatments.

Acknowledgments

The authors would like to thank Ingrid Scholtes and Gill Ison for their help in obtaining the digital pain drawings used in the 2 data sets and Alasdair Kerslake for testing the Body Pain Map tool. They would also like to acknowledge Folkert Tangerman for his feedback regarding the algorithm. AD is presently supported by the UK Medical Research Council and Versus Arthritis Grant (MR/W002426/1) and Advanced Pain Discovery platform. Clinical data sets (data sets 1 and 2) were acquired from separate projects funded by the UK Medical Research Council (MR/J013129/1) and Versus Arthritis (grant 21538).

Conflicts of Interest

None declared.

References

1. Woolf CJ. What is this thing called pain? J Clin Invest 2010 Nov;120(11):3742-3744 [FREE Full text] [doi: 10.1172/JCI45178] [Medline: 21041955]
2. Poquet N, Lin C. The brief pain inventory (BPI). J Physiother 2016 Jan;62(1):52 [FREE Full text] [doi: 10.1016/j.jphys.2015.07.001] [Medline: 26303366]
3. Palmer H. Pain charts; a description of a technique whereby functional pain may be diagnosed from organic pain. N Z Med J 1949 Apr;48(264):187-213. [Medline: 18138692]
4. Jang A, MacLean D, Heer J. BodyDiagrams: improving communication of pain symptoms through drawing. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2014 Presented at: CHI '14: CHI Conference on Human Factors in Computing Systems; Apr 26 - May 1, 2014; Toronto Ontario Canada. [doi: 10.1145/2556288.2557223]
5. Shaballout N, Neubert T, Boudreau S, Beissner F. From paper to digital applications of the pain drawing: systematic review of methodological milestones. JMIR Mhealth Uhealth 2019 Sep 05;7(9):e14569 [FREE Full text] [doi: 10.2196/14569] [Medline: 31489841]
6. Ali SM, Lau WJ, McBeth J, Dixon WG, van der Veer SN. Digital manikins to self-report pain on a smartphone: a systematic review of mobile apps. Eur J Pain 2021 Feb;25(2):327-338 [FREE Full text] [doi: 10.1002/ejp.1688] [Medline: 33113241]
7. Scherrer KH, Ziadni MS, Kong J, Sturgeon JA, Salmasi V, Hong J, et al. Development and validation of the collaborative health outcomes information registry body map. Pain Rep 2021;6(1):e880 [FREE Full text] [doi: 10.1097/PR9.0000000000000880] [Medline: 33490848]
8. Hannover body template. Figshare. URL: https://figshare.com/articles/dataset/Hannover_Body_Template/7637387/9 [accessed 2022-04-26]
9. Kaciroti N, DosSantos MF, Moura B, Bellile EL, Nascimento TD, Maslowski E, et al. Sensory-discriminative three-dimensional body pain mobile app measures versus traditional pain measurement with a visual analog scale: validation study. JMIR Mhealth Uhealth 2020 Aug 19;8(8):e17754 [FREE Full text] [doi: 10.2196/17754] [Medline: 32124732]
10. van den Ameele J, Fuge J, Pitceathly RD, Berry S, McIntyre Z, Hanna MG, et al. Chronic pain is common in mitochondrial disease. Neuromuscul Disord 2020 May;30(5):413-419 [FREE Full text] [doi: 10.1016/j.nmd.2020.02.017] [Medline: 32334903]
11. El-Damanawi R, Lee M, Harris T, Cowley LB, Scholtes I, Bond S, et al. Developing a patient-centred tool for pain measurement and evaluation in autosomal dominant polycystic kidney disease. Clin Kidney J 2021 Nov;14(11):2338-2348 [FREE Full text] [doi: 10.1093/ckj/sfaa259] [Medline: 34754429]

12. Nicholas M, Vlaeyen JW, Rief W, Barke A, Aziz Q, Benoliel R, IASP Taskforce for the Classification of Chronic Pain. The IASP classification of chronic pain for ICD-11: chronic primary pain. Pain 2019 Jan;160(1):28-37. [doi: 10.1097/j.pain.0000000000001390] [Medline: 30586068]
13. Brummett C, Bakshi R, Goesling J, Leung D, Moser S, Zollars J, et al. Preliminary validation of the Michigan Body Map. Pain 2016 Jun;157(6):1205-1212 [FREE Full text] [doi: 10.1097/j.pain.0000000000000506] [Medline: 26835782]
14. Alter BJ, Anderson NP, Gillman AG, Yin Q, Jeong J, Wasan AD. Hierarchical clustering by patient-reported pain distribution alone identifies distinct chronic pain subgroups differing by pain intensity, quality, and clinical outcomes. PLoS One 2021 Aug 4;16(8):e0254862 [FREE Full text] [doi: 10.1371/journal.pone.0254862] [Medline: 34347793]
15. van der Veer SN, Beukenhorst AL, Ali SM, James B, Silva P, McBeth J, et al. Development of a mobile digital manikin to measure pain location and intensity. Stud Health Technol Inform 2020 Jun 16;270:946-950. [doi: 10.3233/SHTI200301] [Medline: 32570521]
16. Rosner J, Lütolf R, Hostettler P, Villiger M, Clijsen R, Hohenauer E, et al. Assessment of neuropathic pain after spinal cord injury using quantitative pain drawings. Spinal Cord 2021 May 16;59(5):529-537 [FREE Full text] [doi: 10.1038/s41393-021-00616-6] [Medline: 33594250]
17. Cruder C, Falla D, Mangili F, Azzimonti L, Araújo LS, Williamon A, et al. Profiling the location and extent of musicians' pain using digital pain drawings. Pain Pract 2018 Jan;18(1):53-66 [FREE Full text] [doi: 10.1111/papr.12581] [Medline: 28466572]
18. Barbero M, Moresi F, Leoni D, Gatti R, Egloff M, Falla D. Test-retest reliability of pain extent and pain location using a novel method for pain drawing analysis. Eur J Pain 2015 Sep;19(8):1129-1138. [doi: 10.1002/ejp.636] [Medline: 25565607]
19. Beissner F, Brünner F, Fink M, Meissner K, Kaptchuk TJ, Napadow V. Placebo-induced somatic sensations: a multi-modal study of three different placebo interventions. PLoS One 2015 Apr 22;10(4):e0124808. [doi: 10.1371/journal.pone.0124808] [Medline: 25901350]
20. Inkscape 1.2. Inkscape. URL: https://inkscape.org/ [accessed 2022-04-26]
21. Body pain map. University of Cambridge. URL: https://pain.medschl.cam.ac.uk/body-pain-map/preview/61c797000aff0789b65d01e6 [accessed 2022-04-26]
22. Python 3.9.0. Python. URL: https://www.python.org/downloads/release/python-390/ [accessed 2022-04-26]
23. Bernard Healey SA, Scholtes I, Abrahams M, McNaughton PA, Menon DK, Lee MC. Role of hyperpolarization-activated cyclic nucleotide-gated ion channels in neuropathic pain: a proof-of-concept study of ivabradine in patients with chronic peripheral neuropathic pain. Pain Rep 2021 Nov;6(4):e967 [FREE Full text] [doi: 10.1097/PR9.0000000000000967] [Medline: 34712888]
24. RS Fibro v1.0. NHS Health Research Authority. URL: https://www.hra.nhs.uk/planning-and-improving-research/application-summaries/research-summaries/rs-fibro-v10/ [accessed 2022-05-03]
25. The brief pain inventory user guide. BPI User Guide. URL: https://www.mdanderson.org/documents/Departments-and-Divisions/Symptom-Research/BPI_UserGuide.pdf [accessed 2022-04-26]
26. Caseiro M, Woznowski-Vu A, De Oliveira AS, Reis FJ, Wideman TH. From paper to digitalized body map: a reliability study of the pain area. Pain Pract 2019 Jul;19(6):602-608. [doi: 10.1111/papr.12780] [Medline: 30884135]
27. emcramer / CHOIRBM. GitHub. URL: https://github.com/emcramer/CHOIRBM [accessed 2022-04-27]
28. Neubert T, Dusch M, Karst M, Beissner F. Designing a tablet-based software app for mapping bodily symptoms: usability evaluation and reproducibility analysis. JMIR Mhealth Uhealth 2018 May 30;6(5):e127. [doi: 10.2196/mhealth.8409] [Medline: 29848470]
29. Ellingsen D, Beissner F, Moher Alsady T, Lazaridou A, Paschali M, Berry M, et al. A picture is worth a thousand words: linking fibromyalgia pain widespreadness from digital pain drawings with pain catastrophizing and brain cross-network connectivity. Pain 2021 May 01;162(5):1352-1363 [FREE Full text] [doi: 10.1097/j.pain.0000000000002134] [Medline: 33230008]
30. NIfTI: Neuroimaging Informatics Technology Initiative. National Institute of Health. URL: https://nifti.nimh.nih.gov/ [accessed 2022-05-08]
31. FMRIB Software Library v6.0. FSL. URL: https://fsl.fmrib.ox.ac.uk/fsl/fslwiki [accessed 2022-04-26]
32. Schott GD. The cartography of pain: the evolving contribution of pain maps. Eur J Pain 2010 Sep;14(8):784-791. [doi: 10.1016/j.ejpain.2009.12.005] [Medline: 20079670]
33. Body mapping in research. In: Handbook of Research Methods in Health Social Sciences. Singapore: Springer; 2019.

Abbreviations

CHOIR: Collaborative Health Outcomes Information Registry
FMRIB: Functional Magnetic Resonance Imaging of the Brain
SSL: Secure Socket Layer
SVG: Scalable Vector Graphics

Edited by A Mavragani; submitted 20.01.22; peer-reviewed by F Beissner, D Wong; comments to author 27.04.22; revised version received 12.05.22; accepted 16.05.22; published 24.06.22.

Please cite as:
Dixit A, Lee M
Quantification of Digital Body Maps for Pain: Development and Application of an Algorithm for Generating Pain Frequency Maps
JMIR Form Res 2022;6(6):e36687
URL: https://formative.jmir.org/2022/6/e36687
doi:10.2196/36687
PMID:35749160

Original Paper

Encouraging Hearing Loss Prevention in Music Listeners Using Personalized Technology: Questionnaire Study

Dylan Tianyu Zhu[1]

Blair Academy, Blairstown, NJ, United States

Corresponding Author:
Dylan Tianyu Zhu
Blair Academy
2 Park St.
Blairstown, NJ, 07825
United States
Phone: 1 9089777453
Email: dylan.zhu.research@gmail.com

Abstract

Background: Noise-induced hearing loss (NIHL) affects millions of people despite being almost completely preventable. For recreational music listening through personal listening equipment (such as earbuds), it seems that listeners do not yet have a way to accurately assess their risk of developing hearing loss and prevent it accordingly.

Objective: The aim of this study is to analyze the perceived utility of a hypothetical device that encourages NIHL prevention based on listeners' exposure to noise and to determine the most effective methods of such encouragement. Here, we describe 3 different potential NIHL risk notification method types, as follows: auditory, external visual, and visual.

Methods: An open, web-based survey was created on Google Forms, and the link was posted to Amazon's Mechanical Turk as well as music-related Reddit communities. The survey was designed to gauge each respondent's self-assessed NIHL awareness, willingness to lower their audio if reminded, and NIHL risk notification type preference. The likelihood of a specific notification type to encourage NIHL prevention among its users was based on the average of each user's responses to 2 survey questions. Data collection started on July 13, 2020, and ended on July 17, 2020.

Results: Of the 116 respondents, 92 (79.3%) reported having prior awareness about NIHL; however, 60 (51.7%) described doing nothing to prevent it despite 96 (82.8%) feeling a moderate, high, or extreme risk of developing NIHL. Of those who already prevented NIHL, 96% (53.5/56) described using estimates to guide their prevention instead of using data. A Kruskal-Wallis test corrected for ties showed that despite the visual NIHL risk notification type being selected by the highest number of participants (84/116, 72.4%), the auditory type had a significantly higher (H_1=6.848; P=.03) average percentage likelihood of encouraging NIHL prevention (62%, SD 24%) among the 40 respondents who chose it, with a median likelihood of 56% (95% CI 50%-75%). The visual type's average likelihood was 50% (SD 28.1%), with a median of 50% (95% CI 37.5%-56.3%). Regardless of the NIHL risk notification type, 69% (80/116) of respondents were not opposed to using NIHL risk notifications and lowering their audio volume accordingly.

Conclusions: The hypothetical device detailed here was thought to be useful because most respondents (82.8%, 96/116) felt an extreme to moderate risk of developing NIHL and such a device could provide accurate data to those who currently use estimates to prevent NIHL, and most respondents were willing to act on NIHL risk notifications. The most effective NIHL risk notification type seemed to be the auditory type, but many aspects of this study need further research to determine which implementation method should reach the public.

(JMIR Form Res 2022;6(6):e24903) doi:10.2196/24903

KEYWORDS

mHealth; mobile health; prevention; NIHL; noise induced hearing loss; MIHL; music induced hearing loss; intervention; wearable device

Introduction

Background

Listening to music is a common recreational activity that occurs in a variety of settings. Although music is often played at a safe volume, if it is played at a sufficiently high volume for a long enough period, it can result in permanent noise-induced hearing loss (NIHL) [1]. An estimated 12.5% of children and adolescents aged 6-19 years (approximately 5.2 million in the United States) and 14% of adults aged 20-69 years (approximately 27.7 million in the United States) have NIHL [2,3]. From 1988 to 2010, Su and Chan [4] did not find a significant increase or decrease in the prevalence of NIHL in adolescents aged 12-19 years, whereas there was a consistently large percentage of the population with NIHL. Across that time frame, the incidence of NIHL in adolescents of that age seemed to be between a lower bound of 13% in the National Health and Nutrition Examination Survey 2009-2010 and an upper bound of 27% in National Health and Nutrition Examination Survey 2007-2008 [4]. Su and Chan [4] also mentioned that adolescent exposure to loud music through headphones increased whereas the use of hearing protection declined. As a nationally representative data set, estimates from these results indicate a large number of NIHL cases among adolescents in the United States. The already problematic incidence rate of NIHL is exacerbated by the fact that NIHL is both irreversible and largely preventable [5,6].

Occupational and Recreational NIHL Differences

NIHL can arise from noise in both occupational and recreational settings [5,7]. Recreational hearing loss is sometimes referred to as music-induced hearing loss, a more specific version of NIHL. As both terms refer to hearing loss caused by noise, NIHL will be the term used in this study. With the rise of personal listening equipment and a general lack of education on NIHL prevention, NIHL is poised to remain a large public health issue in the future [4,8]. Although in the past NIHL was mainly considered to be caused by occupational noise, an increasing number of studies have started citing leisure time sounds as a significant contributor to the development of NIHL [6].

Recreational hearing loss is different from other types of NIHL, which can make it harder to combat. Whereas employers must adhere to regulations for acceptable occupational noise exposure set by the National Institute for Occupational Safety and Health (NIOSH) or the International Organization for Standardization, individual music listeners have complete freedom to adjust the noise entering their ear. If listeners are unaware of the safe noise exposure limits, they can easily exceed them [8]. NIOSH deems a permissible level of noise exposure as noise equal to 8 hours of an equivalent continuous sound level of 85 dB with an exchange rate of 3 dB [9]. Without the equipment to measure and calculate one's incurred noise exposure, especially factoring in the 3-dB exchange rate, consumers can rely only on estimation. As Mercier and Hohmann [10] found that 60% of attendants at a music event did not perceive potentially dangerous audio levels (>87 dB) to be too loud and that 71% already suffered some degree of tinnitus, it seems likely that consumers do not already have the ability to effectively prevent NIHL on their own.

Another aspect of the problem is that a listener's enjoyment of their music can be tied to their music's volume [5]. Mercier and Hohmann [10] found that many young people believe that music is enhanced when played loudly. Kageyama [11] also found that among a group of 46 participants whose median age was 18 years, the sound levels they usually listened to were significantly higher than the levels at which they were comfortable hearing, suggesting that adolescents valued the volume of their music over their personal comfort and possibly the health of their ears. Thus, even if an individual is aware of the risks of NIHL, they may not have the incentive or necessary data to effectively prevent the development of NIHL [5]. Other studies indicate that even when individuals are aware of their risk of developing NIHL, they are still reluctant to use hearing protection [12,13].

Proposed Solution

As there is strong evidence that NIHL can accumulate from a variety of sources [14], care should be taken to reduce NIHL from controllable sources. Preventing recreational hearing loss early is also paramount as it could limit the possibility of incurring greater total hearing loss over the course of one's lifetime [14].

Once NIHL is obtained, the effects of NIHL could perennially cause great emotional, financial, and social stress that could otherwise be avoided [6]. Beyond the more apparent negative effects that NIHL can have on one's quality of life, such as not being able to understand daily conversation, NIHL can also affect the income one receives [15,16]. Neitzel et al [17] found that *considerably conservative* estimates of the economic benefit of preventing NIHL could be anywhere from US $58 billion to US $152 billion in higher salaries.

Kaplan-Neeman et al [18] have previously shown the feasibility of using personal listening equipment and mobile apps to calculate and monitor listening habits. Their study found that people often inaccurately estimate their own listening habits. Participant self-reports for both listening levels and volume settings were only moderately correlated (r=0.655 and r=0.624, respectively) with the actual volume settings. Regardless, even a correct estimation of the actual volume setting would not necessarily mean that the participant had an awareness of their risk of developing NIHL. A total of 22% (8/37) of the participants surpassed the NIOSH daily noise exposure guideline at least once during the 14-day monitoring period.

Existing Solutions

Unlike other potential contexts for music such as concerts, which already have products to mitigate NIHL risk (eg, earplugs), existing products for recreational music listening do not seem detailed enough to be useful to consumers. Apple does allow iPhone operating system (iOS; Apple Inc) 14 users to monitor the output decibel levels of their AirPods through what is called *control center*; however, the control center will not actively notify users about their listening without the user checking it first. In addition, the warning does not factor in the personal listening equipment's noise output over a period, which is an

important aspect of NIOSH's regulation. The warning also does not seem to provide much more discrimination of risk to the user than simply always limiting a certain volume output.

There are many sound meter apps on both the Google Play and Apple App stores, but most only record the noise levels entering the personal listening equipment's microphones or the microphones on the mobile devices themselves, as opposed to recording the output audio of the personal listening equipment itself. Different personal listening equipment may also output the same audio signal differently, which will need to be accounted for on a case-by-case basis. In addition, of the iOS apps we checked which measure the output audio of the earbud itself, many only support newer system versions such as Apple's Health app for iOS 13+, which may limit the number of consumers that could easily make use of the product. A future review of such existing technologies will help better examine the utility of the device discussed in this paper.

Study Objectives

This study seeks to build on the recommendation by Kaplan-Neeman et al [18] to create a technological solution for consumers to objectively monitor their listening habits and better encourage individual NIHL prevention. Providing listeners with some reminder about their music volume based on actual data (rather than estimation) can provide them with the information they need to be more cautious about their own chronic loud noise exposure, especially in response to the sensation-seeking tendency often associated with loud noise [11].

This study is not a randomized controlled trial, as no intervention was applied. It was only a survey. The survey questions were designed to gauge each respondent's self-reported understanding of NIHL and to determine which hypothetical NIHL risk notification method respondents would respond most effectively to. In addition to a phone app NIHL risk notification similar to the method described by Kaplan-Neeman et al [18], 2 additional methods were considered here: auditory notifications (eg, a voice played through the user's personal listening equipment) and external visual notifications (eg, some external device that is not a mobile phone, such as a watch). In the survey, the use case for each NIHL risk notification was described to respondents as being limited to situations where they were listening to music through personal listening equipment such as earbuds or headphones.

It was hypothesized that a significant majority of the population that uses personal listening equipment to listen to music is not aware of NIHL, do not have accurate methods to proactively analyze and prevent NIHL on their own, and are willing to reduce their music volume if reminded to do so. Thus, it was also hypothesized that technology that informs users of their risk of developing NIHL would be useful.

Methods

Survey Design

An open, web-based survey was created on Google Forms with 11 total questions, of which 3 were adaptive follow-up questions. The complete survey can be found in Multimedia Appendix 1. There were 5 pages of questions, of which 2 housed the adaptive follow-up questions (located on pages 2 and 4). The number of questionnaire items per page ranged from 5 to 1. Participants were able to return and change their answers before their survey submission. Before final form submission, they were not given a complete summary of their choices; however, they were able to request an automated summary of their answers after submission. The question order was not randomized among survey respondents. Only for a couple of questions were respondents able to provide a nonresponse answer, which took the form of a free response answer option. The data from the survey were automatically captured in a Google spreadsheet linked to the form. A statement was included at the beginning of the survey to ensure that respondents understood the privacy of the responses they provided, the purpose of the study, and all participants confirmed that they read the privacy statement and consented to having their responses used for this study. Participants were not told about how long or where the data would be stored, nor were they told who the investigator was. As this study was not connected to any institution, we did not apply for approval by an institutional review board.

After asking about the respondent's age, the following three questions were aimed at understanding each survey taker's familiarity with NIHL:

1. "Before this survey, had you heard of noise-induced hearing loss?"
2. "Do you believe you are at risk for developing noise-induced hearing loss?"
3. "Have you been diagnosed with noise-induced hearing loss?"

If a respondent answered "Yes" to the question, "Have you been diagnosed with NIHL?" they were then asked to describe the cause or causes of their NIHL. If a respondent answered "No," they moved on to the next section. The following four questions sought to understand each participant's willingness to mitigate their risk of NIHL:

1. "Do you do anything to actively prevent noise-induced hearing loss?"
2. "Would you like to receive some notification about your risk of developing noise-induced hearing loss from music or the environment around you?"
3. "Would you lower the volume of audio entering your ear if you were reminded?"
4. "Please select all of the notification methods that you would prefer to receive."

If a respondent answered "Yes" to the question, "Do you do anything to actively prevent NIHL?" they were additionally asked if they used quantitative techniques to prevent NIHL. They were also asked to describe their NIHL prevention methods. The 2 questions that asked if the participant would like to receive NIHL risk notifications and if they would be willing to lower their audio volume both used the same 5-point Likert scale. The 5 answer choices on that scale spanned a range of qualitative responses, as follows: definitely, probably, maybe, probably not, and definitely not. The question about NIHL risk notification preferences was set up so that participants were able to select as many notification types as they preferred,

meaning that respondents could each cast votes for more than one method. The available choices were audio, external visual, and visual notifications on your phone, as well as 2 additional choices, other and none. At the end of the survey, a final question was included simply asking respondents to write anything extra they might like to include. For workers on Amazon's Mechanical Turk (MTurk), an additional question asked for their worker ID, and their survey completion screen also included a survey code which workers were asked to paste into MTurk to check for attentive survey completion. The survey was administered in two separate phases: the first phase was to collect responses from MTurk workers, and the second was for responses from Reddit users (Redditors).

Collecting Data Using Amazon's MTurk

MTurk is a website that allows people to be able to answer small tasks. The only requirement for workers on the site is to be aged ≥18 years. MTurk workers neither evidently share a common ideology nor are grouped in some easily discernible way compared with groups on Reddit.

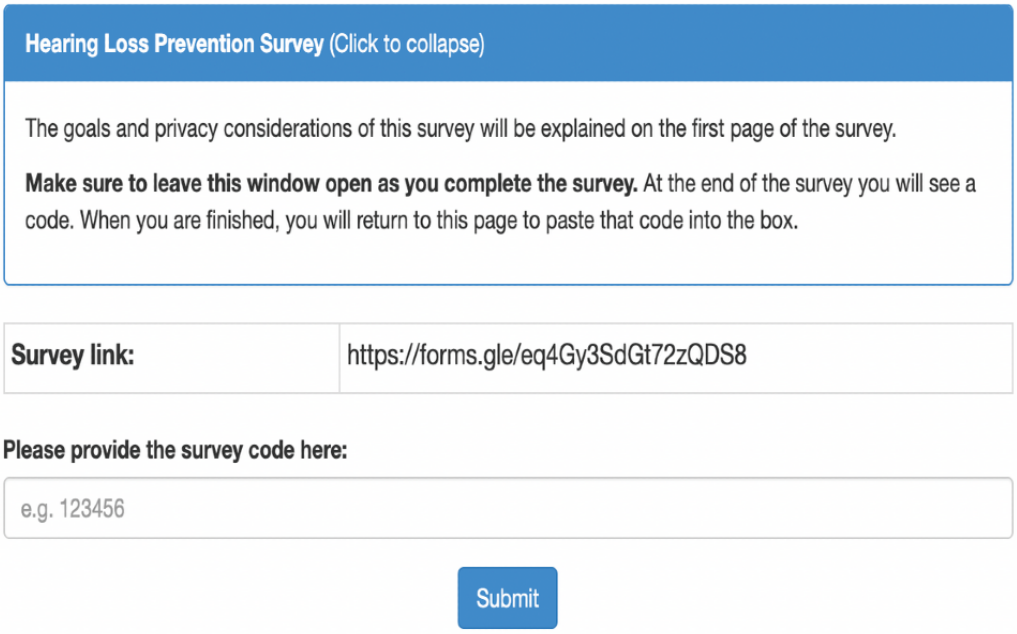
An Amazon MTurk batch was created using the "survey link" template. A picture of what MTurk workers saw can be found in Figure 1. After determining the average survey completion time using a previous MTurk batch, workers were allotted 4 minutes to complete the survey and were compensated with US $0.50 each. Participants were given an estimate of the length of time of the survey in the job description posting on MTurk. The job was not mandatory and all workers chose to take the task from their end. After each batch, worker qualifications (which function like digital tags on MTurk workers) were assigned to every worker ID that had taken the survey. Workers with these qualifications were then filtered out from future survey batches to prevent workers from retaking the survey.

A total of 5 batches were run to collect a total of 195 responses, with the longest batch taking 12 hours to complete and the largest batch requiring 100 responses. MTurk batches were run from July 13, 2020, to July 15, 2020. A summary of all the collected data can be found in Multimedia Appendix 2.

Completed work was rejected if the answer to the adaptive question "How do you prevent NIHL?" was incomprehensible or otherwise unusable for analysis (eg, "yes, i over time of period t0 actively NHL [*sic*]") or meaningless as an answer to the question (eg, "NO"). Such answers reflected a carelessness in answering the survey, which likely carried over to the responses that respondents gave for other questions. As it was otherwise impossible to accurately identify which specific questions had been answered attentively, all their responses were deemed useless for analysis.

As the adaptive question "How do you prevent NIHL?" would only appear to respondents who answered that they do actively prevent NIHL, this response rejection method could not determine the attentiveness of respondents who did not actively prevent NIHL. Thus, if this removal method was continued and the data were then used, the resulting data set would have an artificially high percentage of people who did not already prevent NIHL. However, the data set could not include data filled out by inattentive respondents because of the potential to include false data that do not represent each respondent's true beliefs. As the data points could neither be kept nor correctly removed without introducing unfair bias into the entire data set, all MTurk data were not used for any conclusive analysis. Thus, this study cannot be applied to the general public.

Figure 1. Example of a job request posted to Mechanical Turk.



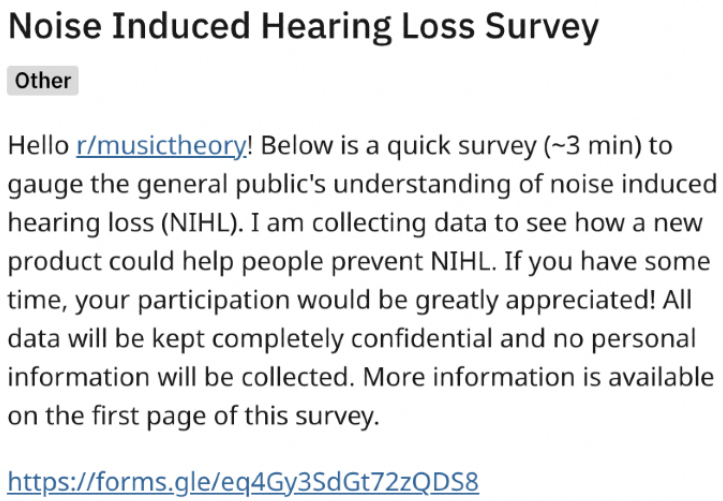
Collecting Data Using Reddit

Reddit was chosen as another site to gather data because it was the easiest and cheapest way to reach a group of users based on a common interest, in this case music. Users on the site often engage in discussions around the theme of each community; these communities are called *subreddits*. The common interest in that theme necessarily preselected participants to be those who enjoyed music to some degree. The goal was to gauge the opinions of recreational music listeners who might use the theoretical product described in this study.

After all MTurk batches had finished, the original survey was slightly modified to remove MTurk-specific features, such as the question about a worker ID and the survey code at the end of the survey. A Reddit post was then drafted with the survey link included, and participants were given an estimate for the survey completion time in the post's description on Reddit. A picture of what each Reddit user saw can be found in Figure 2. The survey was not mandatory. Users self-selected to take the survey from their end and were not actively selected by researchers.

The survey was posted after confirming the approval of each subreddit's moderators. In the end, the subreddits to which the survey was posted were /r/samplesize, /r/WeAreTheMusicMakers, /r/TameImpala, /r/indieheads, /r/musictheory, /r/deathgrips, and /r/Music. The survey remained open to responses for 56 hours starting on July 15, 2020. No incentives were provided to the respondents, financial or otherwise. We did not screen the survey data for repeat respondents, assign cookies to users, or try to determine participation rates. Similarly, IP addresses and log files were not collected or used.

Figure 2. Example of a survey announcement post Reddit, as posted to the subreddit/r/musictheory.



Cleaning Survey Data

As every question on the survey was mandatory in order to submit the form, completion of the survey was implicit in the survey's submission. No survey that was terminated before completion was received. After removing the responses from MTurk workers and 3 Reddit users whose free response answers could not be used (eg, providing answers such as "Poop based fart receptacle," "Death grips," and "poop"), 116 responses made up the final pool of respondent data. Every response gathered from the survey can be found in the Microsoft Excel sheet located in Multimedia Appendix 2. Thus, the resulting sample was a convenience sample of music listeners from Reddit. An original, unmodified copy of the data was preserved on a spreadsheet page in the Microsoft Excel file separate from where data were manipulated to create figures.

Some assumptions were made about certain free responses to create more uniformity in the data (eg, changing the response "all ok" into being a vote for every notification type). In response to the question asking about NIHL prevention methods, 2 respondents used the additional selection choice "Other" instead of either "Exact Decibel Level" or "Estimate" to write a more detailed description of their NIHL prevention method. These responses were sorted into either the "Estimate" or "Exact" groups. The response "i do not i just try to avoid loud place [*sic*]" was counted in the "Estimate" group, and the response "a mix of the two" was counted in both the "Estimate" and "Exact" groups as half of a vote to maintain a consistent total number of respondents (n=116). After every manipulation of the data, the new set was compared with the original to ensure that all information was carried over accurately, and no information was lost or changed without reason.

Ethical Considerations

This study is exempt from a Research Ethics Board review as the methodology falls under the criteria outlined in section 46.104 Exempt research(d)(2)(i) and (d)(2)(i)(ii) of the US Department of Health and Human Services, Basic HHS Policy for Protection of Human Research Subjects [19].

Results

Demographics

The following analysis stems completely from the questions used in the survey, which were quoted in the *Methods* section of this study and are included in both Multimedia Appendices 1 and 2. After removing unusable responses as detailed in the last methods section, there were a total of 116 respondents to

the survey, of whom all were Reddit users. The composition of this group is as follows: 57.8% (67/116) of respondents were between the ages of 18 and 30 years, 31% (36/116) were <18 years old, 8.6% (10/116) were 31-50 years old, and 2.6% (3/116) were >50 years old.

NIHL Awareness

Among music listeners, there was a high awareness of NIHL. Table 1 shows that 79.3% (92/116) respondents had knowledge of NIHL before taking the survey; 2% (2/90) of respondents reported being officially diagnosed with some form of NIHL, although there were a total of 4 respondents (including those 2 respondents) who described having hearing-related issues such as tinnitus and hyperacusis. This suggests that some portion of the population had undiagnosed hearing loss.

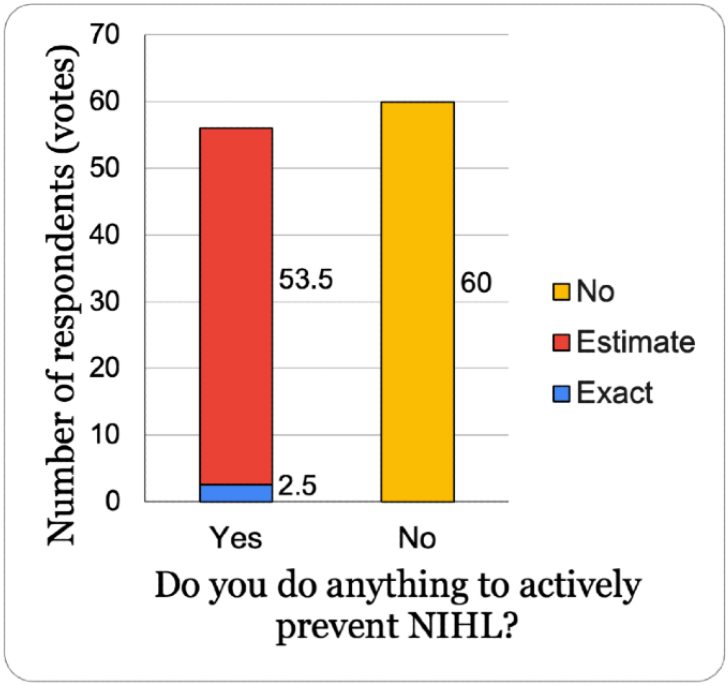
Having more awareness of NIHL seemed to correlate with respondents feeling a greater risk of developing NIHL. Mapping the ordinal responses each to a value from 1 to 5 ("No Risk"=1; "Extreme Risk"=5) showed a slight increase in the mean risk that respondents felt, though an unpaired 2-tailed *t* test showed the difference was not significant (*P*=.16). As shown in Table 1, the median risk felt by all respondents was a moderate risk of developing NIHL. Regardless of prior knowledge of NIHL, about 82.8% (96/116) of all respondents felt a moderate, high, or extreme risk of developing NIHL. Although there was a high awareness of NIHL, Figure 3 shows that a majority of respondents (60/116, 51.7%) did not do anything to prevent NIHL. Of the respondents who did prevent NIHL, 96% (53.5/56) used estimates to prevent NIHL, leaving only about 4% (2.5/56) that claimed to use quantitative data to prevent NIHL.

Table 1. Perceived risk of developing noise-induced hearing loss (NIHL[a]; N=116).

Perceived NIHL risk	Did you have awareness of NIHL before this survey?, n (%)	
	Yes (n=92)	No (n=24)
Extreme risk	5 (5)	0 (0)
High risk	26 (28)	7 (29)
Moderate risk	48 (52)	10 (42)
Little risk	13 (14)	7 (29)
No risk	0 (0)	0 (0)

[a]NIHL: Noise-induced hearing loss.

Figure 3. NIHL awareness and prevention methods. NIHL: noise-induced hearing loss.



NIHL Risk Notification Comparisons

The NIHL risk notification type associated with the greatest likelihood of encouraging NIHL prevention was the auditory type as shown in Table 2.

The percentages shown in Table 2 represent the average likelihood of NIHL prevention across all respondents who voted for a certain NIHL risk notification type. An individual respondent's likelihood of preventing NIHL was determined using their responses to the questions asking if they would like to receive a notification about their risk of developing NIHL, and if they would be willing to lower their media volume if

reminded to. The five possible answer choices for each question were “Definitely,” “Probably,” “Maybe,” “Probably not,” and “Definitely not.” These qualitative responses were mapped to percentage values at regular 25-point intervals, starting with “Definitely not” being equivalent to 0% and ending with “Definitely” as equivalent to 100%. Each respondent’s likelihood of preventing NIHL was then estimated to be the product of their answers to the 2 questions. A table of all possible percentages and their products is presented in Table 3.

The average likelihood of eventual NIHL prevention across all respondents who voted for the auditory NIHL risk notification (n=40) was 62% (SD 24%) with a median of 56% (95% CI 50%-75%). This was significantly higher than the average across all visual NIHL risk notification voters (n=84), which was 50% (SD 28%) with a median of 50% (95% CI 38%-56%). There was no significant difference (H_1=6.365; P=.19) between the auditory NIHL risk notification and the external visual NIHL risk notification, whose average probability of encouraging NIHL was 52% (SD 30%) and had a median of 57% (95% CI 38%-75%). A Kruskal-Wallis test corrected for ties revealed a significant difference (H_1=6.848; P=.03) between the auditory and visual NIHL risk notifications. No significant difference was observed between visual and external visual NIHL risk notifications (H_1=6.307; P=.64) either.

As shown in Table 2, where “n” reflects the number of unique respondents who chose a notification type, the NIHL risk notification that was selected the most often was the visual NIHL risk notification with 84 votes. The auditory and external visual NIHL risk notifications each received 40 votes. A total of 12 respondents chose the “None” selection, with the average chance of preventing NIHL in that group being 14% (SD 21%). A total of 3 respondents chose “Other,” and their mean likelihood of preventing NIHL was 21% (SD 16%). The three “Other” responses provided by respondents are listed below exactly as they were typed by each respondent:

1. “Nonintrusive pop-up in a top empty corner, warning for SERIOUS issues (that can be deactivated) and not bug you about everything, as phones do.”
2. “Vibration maybe?”
3. “Haptic, on a watch or something.”

Table 2. Average likelihood of each noise-induced hearing loss (NIHL[a]) risk notification type to encourage NIHL prevention.

NIHL risk notification type	Likelihood (%), mean (SD)
Auditory (n=40)	62 (24)
External visual (n=40)	52 (29)
Visual (n=84)	50 (28)
Other (n=3)	21 (16)
None (n=12)	14 (21)

[a]NIHL: Noise-induced hearing loss.

Table 3. Response ranking system.

Possible responses	Definitely (%)	Probably (%)	Maybe (%)	Probably not (%)	Definitely not (%)
Definitely	100	75	50	25	0
Probably	75	56.3	38.5	18.8	0
Maybe	50	38.5	25	12.5	0
Probably not	25	18.8	12.5	6.3	0
Definitely not	0	0	0	0	0

Respondent Willingness to Prevent NIHL

As shown in Table 4, on average, most respondents were inclined to lower the volume of audio entering their ears. The percentages listed in Table 4 are related to the total number of respondents for that age range. Regardless of age, 66.4% (77/116) of respondents would either “Definitely” or “Probably” turn down their audio volume if reminded to. In addition, of those respondents who did not already prevent NIHL, 57% (34/60) would either “Definitely” or “Probably” turn down their audio volume. A total of 18.9% (22/116) respondents chose “Maybe,” and 14.7% (17/116) respondents chose either “Probably not” or “Definitely not.”

Table 4. Table of respondent willingness to turn down audio by age group.

Respondents	Would you lower the volume of audio entering your ear if you were reminded to?, n (%)				
	Definitely	Probably	Maybe	Probably not	Definitely not
<18 years old (n=36)	10 (27.8)	13 (36.1)	9 (25)	3 (8.3)	1 (2.8)
18-30 years old (n=67)	22 (32.8)	23 (34.3)	12 (17.9)	10 (14.9)	0 (0)
31-50 years old (n=10)	4 (40)	3 (30)	1 (10)	1 (10)	1 (10)
>51 years old (n=3)	0 (0)	2 (66.7)	0 (0)	1 (33.3)	0 (0)
All respondents (n=116)	36 (31)	41 (35.3)	22 (19)	15 (12.9)	2 (1.7)

Discussion

Principal Findings

Overall, the data seemed to suggest that technology that can provide music listeners with a real-time assessment of their risk of developing NIHL would be useful for encouraging NIHL prevention. A majority (60/116, 51.7%) of the population did not practice any form of NIHL prevention despite 79.3% (92/116) of respondents having some awareness of NIHL. Even within the population that did prevent NIHL, only about 4% (2.5/56) reported using noise level data to aid in the prevention of NIHL. Across the entire population of respondents, 82.8% (96/116) felt more than a little risk of developing NIHL. None of the respondents reported a complete lack of risk. Of the respondents, 66.3% (77/116) were willing to lower their audio volume if they were reminded to, and the NIHL risk notification that had the highest mean estimated chance of encouraging NIHL prevention was the auditory NIHL risk notification at 61.9%. The NIHL risk notification type respondents preferred the most seemed to be the visual NIHL risk notification type with 84 total votes.

As there was a high percentage (53.5/56, 96%) of respondents who used estimates in their prevention of NIHL, and 51.7% (60/116) of all respondents did not practice preventing NIHL, technology that can both encourage NIHL prevention in those who do not already prevent it and provide more accurate data to those who currently estimate their noise exposure would be useful. Without data to guide prevention methods, consumers of music may either not truly prevent NIHL or be listening to music at too low a volume to enjoy the activity. For those who do not already prevent NIHL, technology that encourages NIHL prevention could allow them to enjoy their music for a longer period of their lives. Even though a majority of the population did have awareness of NIHL, technology that informs the user of their risk of developing NIHL could still be useful because most respondents did not prevent it, do not have accurate methods of actively analyzing and preventing NIHL on their own, and are willing to reduce their music volume if reminded to.

Future Work

The results of this study readily lend themselves to many opportunities for future research and overall seem to suggest an encouraging environment for the future of NIHL awareness and prevention. Instead of increasing NIHL awareness through conventional methods such as bolstering NIHL education in schools, future efforts to increase NIHL awareness and prevention could take the form of more targeted individual encouragement through technology. Below are various flaws in the research methodology that seriously affected the conclusions that could be drawn, which future studies would do well to improve upon.

Study Population

In choosing Reddit communities already related to music, the survey collection method likely encouraged a bias toward respondents who heavily listened to music, and thus, may disproportionately favor using an NIHL risk notification compared with the general public. This study may also not have fully sampled a group that was representative of people who regularly listen to music, as those music listeners who were not in the chosen subreddits were not given the same opportunity to supply their responses. This bias may have resulted in a seemingly greater concern for preserving hearing ability than may truly reflect the population of people who would benefit from awareness about NIHL. In addition, without an external incentive to complete the survey, respondents who did go out of their way to supply responses might have already believed more in the utility of an NIHL risk notification. To obtain a more accurate estimate of the general public's perception of NIHL risk notifications, future surveys should use a population more representative of the general public and provide an additional incentive to complete the survey, such as a monetary reward. A more randomized sampling method should also be used to create a sample of those who enjoy music. Future surveys should also try to include more participants aged ≥31 years, as well as possibly control for the countries of each participant. The timing of the survey's posting may also be important to for control in future studies.

Data Collection

A more quantitative data collection method to determine the usefulness of each NIHL risk notification should be used in the future. The Kruskal-Wallis test used here inflated the possibility of type 1 errors, especially with having to correct for ties in the data. Future studies could also incorporate data from pure tone audiometry measurements and otoacoustic emission tests. There should also be a way to assess the true level of NIHL knowledge each participant had, instead of letting them self-report either a "Yes" or a "No." It may also be easier to gauge how respondents feel toward individual NIHL risk notification types by having a more direct rating system for each individual NIHL risk notification. A picture of each hypothetical NIHL risk

notification would also have made each notification type easier to conceptualize for respondents.

MTurk Data

Although the MTurk data may not truly reflect a more general population of people, some apparent differences between the Reddit and MTurk data suggest that NIHL risk notifications might not be as useful for the general public. A much greater percentage (35/63, 56%) of MTurk workers than Reddit users (2.5/56, 4.5%) indicated that they used exact decibel levels to inform their NIHL prevention. This seems counterintuitive because conventional logic might suggest that people who interact with audio more would be more likely to have the technology and inclination to accurately track their risk. In addition, a higher percentage of MTurk respondents (54/121, 44.6%) than Reddit respondents (12/116, 10.3%) indicated that none of the listed notification types were good, suggesting that a different notification type may be necessary to encourage NIHL prevention in the general public. Interestingly, a much higher percentage (55/121, 45.5%) of MTurk respondents indicated that they had officially diagnosed NIHL compared with 1.7% (2/116) of respondents in the Reddit population. It should be noted again that the MTurk data can only indicate a need to more accurately study the general public to either confirm or reject these trends, as the various flaws in the response collection process delegitimize the significance of any conclusions that could be drawn. Many MTurk responses also suggested that respondents did not understand English very well, which may have led to them unintentionally answering questions inaccurately.

Changing Behaviors

Despite this survey describing some likelihood of NIHL risk notifications (especially the auditory type) successfully encouraging NIHL prevention, significantly more research needs to be done to determine which NIHL risk notification could actually encourage lasting behavioral change in practice. Within each of the three NIHL risk notification types identified here, there are numerous design choices to be considered during the production of actual prototypes, which may change the overall effectiveness of each NIHL risk notification. Choices such as using more images over text in a visual notification or perhaps using different people's voices for the audio NIHL risk notification all may need to be considered. Similar to the work by Kaplan-Neeman et al [18], future work must be done with all forms of NIHL risk notifications to conclusively determine what system and data presentation methods might actually encourage user change.

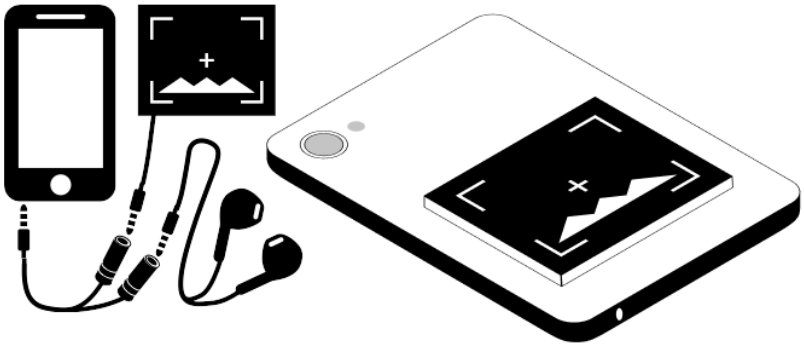
Moving Forward

The three NIHL risk notifications described here correlate to 3 distinct types of consumer products that could be created in the future. External visual NIHL risk notifications could take the form of wearable watches, as mentioned earlier, or something similar to Figure 4, where a physical connection to the user's media source is maintained. "Audio Jack" and "screen" icon sources from the study by Daksina [20] and Ahmad [21], respectively. The overall figure was designed by author DTZ. It should be noted that a potential flaw in the utility of this risk notification type may be the current consumer trend toward wireless listening equipment. The audio NIHL risk notification was conceptualized as a software and hardware package that both performed the required noise exposure calculations and provided the NIHL risk notifications as audio cues all from within the user's personal listening equipment. The visual NIHL risk notification was mainly envisioned as an app on the user's phone to measure the audio equipment output and then visually display NIHL warnings to the user.

Another NIHL risk notification method was later theorized by Tim Nieuwenhuis, which adds to the potential utility of the visual notification method. In the idea, an external physical device would act as the user's ear and measure the noise output of the personal listening device as the user plays a provided audio file for calibration. The device would then send the data back to the user's phone. The NIHL risk notification app combines the output data with the audio signal data going out of the phone to build a model for estimating the noise that enters the user's ear. The notification itself would likely still appear on the app. There is potential for this method to be applicable to recreational music listening through speakers, although an issue then might be determining the user's position in a room.

It may also be worthwhile to adapt NIHL risk notifications for contexts beyond recreational music listening at home. Future NIHL risk notifications could be created for other situations with a high risk of causing NIHL, such as playing an instrument or attending music festivals [6,22]. A version might be created to augment the usefulness of existing earplugs. Significant consideration should be given to the fact that the NIHL risk notifications theorized here rely heavily on being able to use the user's personal listening equipment to calculate a time-weighted dBA average of noise exposure entering the ear, which may not be as feasible in other contexts or within equipment such as earplugs.

Figure 4. A visual screen is attached to the back of a mobile phone, similar to existing portable digital-to-analog converters. The audio connection is facilitated by a male to 2-female audio splitter cable.



Acknowledgments

The authors would like to thank Professor Lorie Loeb's supportive tutelage, without which this study would not have been possible or even conceived. They would also like to thank Tim Nieuwenhuis for his novel perspective on another potential noise-induced hearing loss risk notification method.

Conflicts of Interest

DTZ has an interest in personally creating the hypothetical device described in this study and believes it could be a useful product. The device described here has not been developed by the author.

Multimedia Appendix 1
The complete Google Form as presented to survey respondents.
[PDF File (Adobe PDF File), 277 KB - formative_v6i6e24903_app1.pdf]

Multimedia Appendix 2
Complete catalog of all survey data collected in this study. The compiled data from every survey batch are on the spreadsheet page titled "Compiled," which has already been sorted. The "Cleaning Data" page was where all the data were manipulated to create graphs. Preliminary data tables can be found on the far-right side of the page.
[XLSX File (Microsoft Excel File), 195 KB - formative_v6i6e24903_app2.xlsx]

References

1. Sliwinska-Kowalska M, Davis A. Noise-induced hearing loss. Noise Health 2012;14(61):274-280. [doi: 10.4103/1463-1741.104893] [Medline: 23257577]
2. Niskar AS, Kieszak SM, Holmes AE, Esteban E, Rubin C, Brody DJ. Estimated prevalence of noise-induced hearing threshold shifts among children 6 to 19 years of age: the Third National Health and Nutrition Examination Survey, 1988-1994, United States. Pediatrics 2001 Jul;108(1):40-43. [doi: 10.1542/peds.108.1.40] [Medline: 11433052]
3. Hoffman HJ, Dobie RA, Losonczy KG, Themann CL, Flamme GA. Declining prevalence of hearing loss in US adults aged 20 to 69 years. JAMA Otolaryngol Head Neck Surg 2017 Mar 01;143(3):274-285 [FREE Full text] [doi: 10.1001/jamaoto.2016.3527] [Medline: 27978564]
4. Su BM, Chan DK. Prevalence of hearing loss in US children and adolescents: findings from NHANES 1988-2010. JAMA Otolaryngol Head Neck Surg 2017 Sep 01;143(9):920-927 [FREE Full text] [doi: 10.1001/jamaoto.2017.0953] [Medline: 28750123]
5. Daniel E. Noise and hearing loss: a review. J Sch Health 2007 May;77(5):225-231. [doi: 10.1111/j.1746-1561.2007.00197.x] [Medline: 17430434]
6. McCormick J, Matusitz J. The impact on U.S. Society of noise-induced and music-induced hearing loss caused by personal media players. Int J Listen 2010 May 06;24(2):125-140. [doi: 10.1080/10904011003744565]
7. Masterson EA, Themann CL, Luckhaupt SE, Li J, Calvert GM. Hearing difficulty and tinnitus among U.S. workers and non-workers in 2007. Am J Ind Med 2016 Apr;59(4):290-300. [doi: 10.1002/ajim.22565] [Medline: 26818136]
8. Zhao F, Manchaiah VK, French D, Price SM. Music exposure and hearing disorders: an overview. Int J Audiol 2010 Jan;49(1):54-64. [doi: 10.3109/14992020903202520] [Medline: 20001447]

9. Schulte PA, Dunn DE. Occupational noise exposure: revised criteria 1998. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention. 1998. URL: https://www.cdc.gov/niosh/docs/98-126/pdfs/98-126.pdf?id=10.26616/NIOSHPUB98126 [accessed 2021-09-16]
10. Mercier V, Hohmann BW. Is electronically amplified music too loud? What do young people think? Noise Health 2002;4(16):47-55. [Medline: 12537841]
11. Kageyama T. Loudness in listening to music with portable headphone stereos. Percept Mot Skills 1999 Apr;88(2):423. [doi: 10.2466/pms.1999.88.2.423] [Medline: 10483631]
12. Bogoch II, House RA, Kudla I. Perceptions about hearing protection and noise-induced hearing loss of attendees of rock concerts. Can J Public Health 2005;96(1):69-72 [FREE Full text] [Medline: 15682701]
13. Widén SE, Erlandsson SI. The influence of socio-economic status on adolescent attitude to social noise and hearing protection. Noise Health 2004;7(25):59-70. [Medline: 15703150]
14. Harrison RV. The prevention of noise induced hearing loss in children. Int J Pediatr 2012;2012:473541 [FREE Full text] [doi: 10.1155/2012/473541] [Medline: 23304173]
15. Ruben RJ. Redefining the survival of the fittest: communication disorders in the 21st century. Laryngoscope 2000 Feb;110(2 Pt 1):241-245. [doi: 10.1097/00005537-200002010-00010] [Medline: 10680923]
16. Shield B. Evaluation of the social and economic costs of hearing impairment: a report for Hear-It Aisbc. Hear-It. 2006. URL: https://www.hear-it.org/sites/default/files/multimedia/documents/Hear_It_Report_October_2006.pdf [accessed 2021-09-16]
17. Neitzel RL, Swinburn TK, Hammer MS, Eisenberg D. Economic impact of hearing loss and reduction of noise-induced hearing loss in the United States. J Speech Lang Hear Res 2017 Jan;60(1):182-189. [doi: 10.1044/2016_jslhr-h-15-0365]
18. Kaplan-Neeman R, Muchnik C, Amir N. Listening to music with personal listening devices: monitoring the noise dose using a smartphone application. Int J Audiol 2017 Jun 10;56(6):400-407. [doi: 10.1080/14992027.2017.1297541] [Medline: 28281836]
19. US Department of Health & Human Services. Basic HHS Policy for Protection of Human Research Subjects. Title 45: Public Welfare, Department of Health and Human Services. Part 46: Protection of Human Subjects. 2018. URL: https://www.hhs.gov/ohrp/regulations-and-policy/regulations/45-cfr-46/revised-common-rule-regulatory-text/index.html#46.104
20. Daksina D. 'Audio Jack' SVG file. The Noun Project. URL: https://thenounproject.com/term/audio-jack/2158719/ [accessed 2021-09-16]
21. Ahmad ID. 'Screen' SVG File. The Noun Project. URL: https://thenounproject.com/term/screen/2090101/ [accessed 2021-09-16]
22. Halevi-Katz DN, Yaakobi E, Putter-Katz H. Exposure to music and noise-induced hearing loss (NIHL) among professional pop/rock/jazz musicians. Noise Health 2015;17(76):158-164 [FREE Full text] [doi: 10.4103/1463-1741.155848] [Medline: 25913555]

Abbreviations

MTurk: Mechanical Turk
NIHL: noise-induced hearing loss
NIOSH: National Institute for Occupational Safety and Health

Edited by G Eysenbach; submitted 19.01.21; peer-reviewed by D Goyal, J Zhang; comments to author 02.03.21; revised version received 16.07.21; accepted 28.07.21; published 27.06.22.

Please cite as:
Zhu DT
Encouraging Hearing Loss Prevention in Music Listeners Using Personalized Technology: Questionnaire Study
JMIR Form Res 2022;6(6):e24903
URL: https://formative.jmir.org/2022/6/e24903
doi:10.2196/24903
PMID:35759318

Original Paper

Feasibility and Acceptability of a Ugandan Telehealth Engagement Platform for Informational Messaging on Modern Contraception: Pilot Cross-sectional Study

Louis Henry Kamulegeya[1], MD, MPH; JohnMark Bwanika[1], MD, MSc; Joy Banonya[1], BSc; Joan Atuhaire[1], BSc, MSc; Davis Musinguzi[1], MD, MBA; Vivian Nakate[2], BSc; Joshua Kyenkya[2], BSc; Lydia Namatende[2], BSc; Keith J Horvath[3], MD, MSc, PhD; Agnes Kiragga[2], BSc, PhD

[1]The Medical Concierge Group, Projects and Research Department, Kampala, Uganda

[2]Infectious Diseases Institute, Makerere University, Department of Research, Kampala, Uganda

[3]Department of Psychology, San Diego State University, San Diego, CA, United States

Corresponding Author:
Louis Henry Kamulegeya, MD, MPH
The Medical Concierge Group
Projects and Research Department
Lugogo Bypass, Mirembe Business Center
Kampala, 256
Uganda
Phone: 256 783216066
Email: louis@tmcg.co.ug

Abstract

Background: With the region's highest population growth rate (30%), Uganda is on the brink of a population explosion, yet access to and utilization of public health control measures like modern contraception is a challenge. This is due to remotely located health facilities, noncustomized health content, and poor or nonfunctional post-facility follow-up.

Objective: The aim of our study was to evaluate the feasibility and acceptability of a telehealth engagement platform primarily targeting men; the platform provided behavioral and informational messaging on modern contraception (ie, family planning) and its impact on shaping sexual and reproductive health and knowledge and uptake of family planning services.

Methods: A longitudinal cohort of men aged 18 years and older gave consent to receive mobile phone messages on family planning; follow-up was performed at months 1, 4, and 6 to assess key study-related outcomes on knowledge transfer and acquisition on modern contraception, partner communication, and spousal uptake of family planning. Qualitative interviews with the study participants' spouses were also performed.

Results: The study included 551 study participants, 450 of whom were men, the primary study participants, who received the family planning mobile messages and 101 of whom were their spouses. Of the 450 primary participants, 426 (95%) successfully received the messages and only 24 (5%) reported not receiving them. The average response (ie, participation) rate in weekly quizzes was 23%. There was a noted 18.1% increase in couple communication attributed to the intervention; couples opened up more to each other on matters concerning family planning.

Conclusions: Using digital channels to address the concerns and inquiries of participants in real time or as fast as possible helped to increase the likelihood that couples adopted family planning.

(JMIR Form Res 2022;6(6):e34424) doi:10.2196/34424

KEYWORDS

telehealth; mHealth; digital health; family planning; contraception; messaging; male involvement; health education; Uganda

Introduction

Background

In 2021, Uganda's population was 42.4 million people, representing growth of about 30% compared to 2014, when the national population census reported a population of 36.9 million [1]. Over 55% of the country's population was below the age of 15 years and thus close to childbearing age.

According to the 2016 Uganda Demographic Health Survey, the total average fertility rate was 5.4 children per woman [2]. However, the rate is known to be higher in rural and semiurban parts of the country, where on average a woman gives birth to 7 children in her lifetime, making Uganda a country with one of the fastest growing populations in the world. These trends point to a looming population explosion in a setting with high poverty levels, low literacy rates, and limited access to quality health services unless population mitigation measures are urgently taken up.

Population explosion control measures, such as modern contraception, have shown promise in driving socioeconomic growth and political stability in sub-Saharan Africa [3,4]. However, modern contraception (ie, family planning) uptake in Uganda had tended to remain suboptimal, with a 30% contraceptive prevalence rate; the current unmet need for family planning among women has been placed at 34% [5,6].

A number of factors have been identified to account for the above trends, including a lack of access to credible information on modern contraception, lack of male partner support and engagement in decision-making, and poor or nonfunctional post-service follow-up mechanisms to address challenges like side effects, myths, and misconceptions [7,8].

Digital Technology and Reproductive Health Services Delivery in Africa

The application of information technology and digital tools in the delivery of sexual and reproductive health (SRH) services is gaining momentum, with the use of artificial intelligence (AI), short message service (SMS) messaging, and hotlines having been documented. For example, askNivi, an AI chatbot, was piloted in Kenya as a demand-generation tool for contraception uptake targeting adolescents and young women; it resulted in a 41% probable increase in the likelihood of contraceptive uptake among users [9]. Similarly, a study by Njagi J [10] showed that helplines (also called hotlines) provided an alternative, reliable channel for young girls and women to seek clarity and guidance on their SRH issues in a society where the nature of adult-child relations is hierarchical and conservative. The use of mobile SMS for the dissemination of health information on family planning and antenatal attendance reminders for pregnant women has been piloted in sub-Saharan Africa and has made a positive impact [11,12].

Goal of This Study

The aim of our study was to evaluate the feasibility and acceptability of a telehealth engagement platform primarily targeting men, with behavioral and informational messaging on modern contraception (ie, family planning) and its impact on shaping their knowledge of sexual and reproductive health (SRH) and their uptake of family planning services.

Methods

Sample Size Calculation

The primary outcome used to estimate the study's sample size was the change in uptake of family planning services of participants who received the men's telehealth information package (mTIP) intervention. According to the most recent Uganda Demographic Health Survey, held in 2016, the proportion of men aged 15 to 55 that use any family planning method was 60% [2]. Assuming that there would be a 25% increase in family planning uptake due to the intervention, 5% type I error, and 90% power, we calculated that we would need to enroll 432 men. The sample size was then adjusted for an anticipated 10% attrition over the 3-month study period, which gave us a total of 475 men as the sample size to be recruited.

Inclusion and Exclusion Criteria

Potential study participants were eligible for participation if they were aged 18 to 55 years, had a spouse or current active sexual partner (sexually active was defined as having at least 1 sexual encounter in the previous 6 months), owned a mobile phone, and were willing to take part in all study-related activities, especially periodic surveys using questionnaires. Study participants that were unable to effectively comprehend the study-related activities, unable to communicate due to suboptimal mental status or low literacy levels, or did not own a phone were excluded.

Study Setting and Participant Recruitment

We targeted men 18 years and older who consented to participate in the study. The participant recruitment took place in 8 community settings that included academic institutions, workplaces, and social gatherings, such as at sports grounds. The study team set up a tent at each site with an appropriate level of privacy and confidentiality for the purposes of the informed consent process. Study participants that consented to take part in the study were then required to send a trigger message to the study's SMS prepaid short code so as to start receiving mobile messages on family planning. The SMS short code was operated and maintained at The Medical Concierge Group (TMCG), a digital health and telemedicine company headquartered in Uganda that was one of the study partners. Prior to full study enrollment, a beta-test study with 25 participants was performed, with the findings used to improve the data collection tools and informed consent documents, which were resubmitted to our institutional review board for approval.

Mobile Message Design and Dissemination

The information-behavior-motivation (IBM) skills model formed the basis of the development of the mobile messages. This model has been utilized in behavioral change approaches with the end goal of influencing adoption of positive behavior through providing correct information for informed decision-making and creating an environment that motivates the adoption of this positive behavior, for example, through reminders and nudges [13,14]. Messages on SRH focusing on modern contraception

were developed by the study team and reviewed by a community advisory board and an SRH specialist for appropriateness, relevance, and local context. The messages covered IBM aspects of contraception and family planning communication. The messages were designed to go out on a weekly schedule via a prepaid short code (8884) with an average of 2 messages received weekly over a period of 60 days by the study participants. The messages were delivered in English.

Study Participant Follow-up and Engagement

All study participants had access to a study toll-free telephone and SMS platform that was available 24 hours a day and staffed by qualified health professionals to offer remote resolutions to participants' inquiries, including referrals and links to SRH and other health services. In addition, proactive follow up was performed by the study team at months 1, 4, and 6 after the date of study enrollment to perform specific study procedures and assess the participants for the knowledge they had gained on modern contraceptive methods, couple communication on family planning, and partner uptake of family planning.

Data Collection

During the scheduled routine follow-up phone calls, the study participants were interviewed by one of the study staff, who was trained in phone interviews, at TMCG. During the interviews, the study team interacted with the participants to assess their awareness of family planning methods, the men's attitudes and practices, self and spousal use of family planning, spousal communication about family planning decision-making, and the men's opinions about their roles in family planning decision-making. The interview dates and times were negotiated and agreed upon by both the study staff and participants. The interviews were conducted in either the local language (Luganda) or English. A pretested electronic questionnaire built on an open data kit was used to collect information on the participants' experiences with the telehealth platform and phone ownership. The study telehealth platforms (ie, SMS and the hotline) were analyzed for performance on message delivery, study participants' engagement in quizzes, and completion of all study requirements. In-person short interviews were conducted with 25 randomly selected study participants (15 men and 10 women) to gather insights on the feasibility and accessibility of the mTIP intervention.

Data Analysis and Interpretation

Quantitative data collected through the open data kit were analyzed using Stata software (StataCorp). Quantitative data were collected through TMCG's telehealth platforms (SMS and the hotline) following the dissemination of the information on family planning via mobile phone. This focused on the number of SMS messages and voice calls, number of referrals, number of participants who completed all study assessments, and any other data regarding family planning, which were summarized and used as a measure of feasibility and scalability. For a qualitative inquiry, the study used in-depth interviews to elicit information from 15 men and 10 women on their experiences regarding family planning. The audio data were transcribed, coded, and thematically analyzed to address the objectives of the study. Multiple data sources from in-depth interviews with both male and female participants were used for data triangulation. We also used an information-rich description of the findings to ensure transparency. Table 1 summarizes the demographic characteristics of the interview participants.

Table 1. Summary of characteristics of the participants who underwent in-depth interviews.

No.	Age, years	Sex	Occupation	Marital status	Children, n	Religion	Level of education
1	56	Male	Unemployed	Married	8	Anglican	University
2	27	Male	Businessperson	Single	0	Catholic	University
3	32	Male	Teacher	Married	3	Born again	University
4	20	Male	Student	Married	0	Born again	University
5	24	Male	Driver	Married	1	Catholic	Primary
6	28	Male	Private employee	Married	2	Muslim	Secondary
7	30	Female	Private employee	Married	2	Born again	Secondary
8	25	Female	Businessperson	Married	1	Born again	Secondary
9	22	Female	Farmer	Married	2	Anglican	Secondary
10	30	Female	Casual laborer	Married	2	Catholic	Primary
11	29	Female	Teacher	Married	2	Catholic	Tertiary
12	29	Female	Housewife	Married	3	Catholic	No school
13	50	Male	Civil servant	Married	0	Seventh-day Adventist	University
14	41	Male	Religious leader	Married	5	Seventh-day Adventist	Tertiary
15	24	Male	Student	Single	0	Seventh-day Adventist	University
16	25	Male	Casual laborer	Married	1	Born again	Secondary
17	30	Male	Self employed	Married	3	Catholic	University
18	38	Male	Police officer	Married	4	Anglican	University
19	30	Male	Teacher	Married	1	Catholic	Tertiary
20	38	Male	Former security	Married	4	Catholic	Primary
21	35	Male	Teacher	Married	4	Catholic	University
22	26	Female	Businessperson	Married	2	Born again	Secondary
23	28	Female	Private employee	Married	0	Anglican	University
24	23	Female	Unemployed	Married	1	Unknown	No school
25	32	Female	Teacher	Married	6	Seventh-day Adventist	Tertiary

Ethics Approval

The study was approved by the Joint Clinical Research Centre institutional review board (approval number 0906-2019) and registered with the Uganda National Council of Science and Technology (reference number HS425ES). All the study procedures, compensation, benefits, potential risk of participation, and the voluntary and confidential nature of participation were discussed. Written informed consent was obtained from all respondents before enrollment in this qualitative study. For young adults with low literacy, we used a thumbprint in the presence of a witness.

Results

Demographic Characteristics

A total of 551 study participants were recruited, including 450 men (the primary study participants) and 101 women (their spouses), who were proactively observed by the study team over a 6-month period via voice follow-up calls at months 1, 4, and 6 after enrollment. The demographic characteristics of the study participants are summarized in Table 2.

Table 2. Demographic characteristics of the study participants.

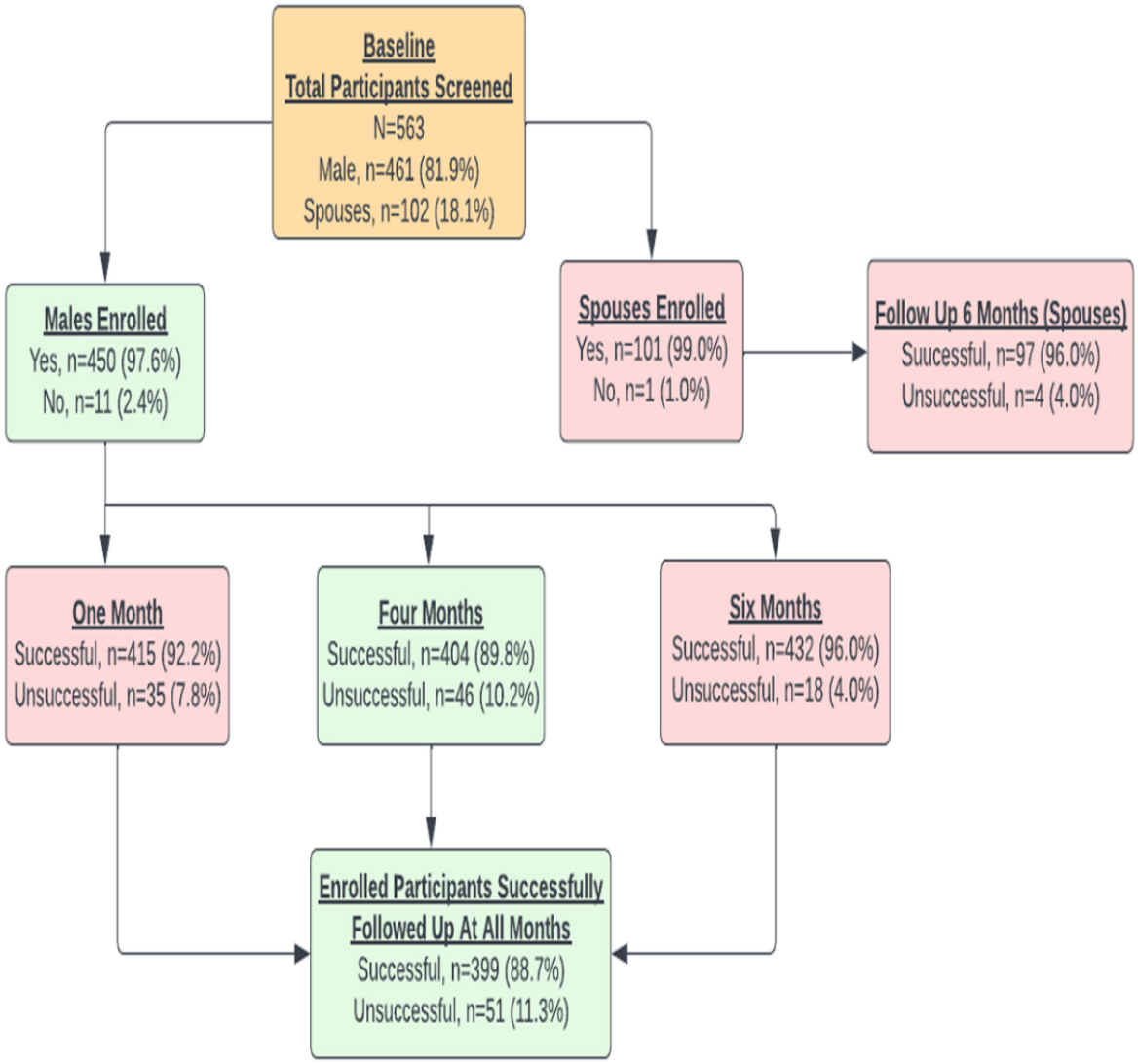
Variable	Men (N=450)	Spouses (N=101)
Age (years), median (IQR)	25 (22-30)	25 (23-28)
Marital status, n (%)		
Single	257 (57.1)	15 (14.9)
Married (religious, civil, or customary)	177 (39.3)	59 (58.4)
Widowed, separated, or divorced	8 (1.8)	17 (16.8)
Has spouse, but not legally married	8 (1.8)	10 (9.9)
Current occupation, n (%)		
Student	143 (31.8)	15 (14.9)
Employed	296 (65.8)	59 (58.4)
Unemployed	7 (1.6)	17 (16.8)
Declined to answer	4 (0.9)	10 (9.9)
Education level, n (%)		
No education (did not complete any education)	2 (0.4)	2 (2)
Primary level	62 (13.8)	9 (8.9)
Secondary level	234 (52)	60 (59.4)
University[a]	149 (33.1)	29 (28.7)
Declined to answer	3 (0.7)	1 (1)
Type of digital device owned (multiple choice), n (%)		
Basic mobile phone	281 (60.8)	39 (36.1)
Smartphone	177 (38.3)	69 (63.9)
Desktop computer	2 (0.4)	0 (0)
Laptop	1 (0.2)	0 (0)
Tablet	1 (0.2)	0 (0)

[a]Makerere and Kyambogo Universities are both non–faith-based tertiary institutions.

User Statistics

The study participants were observed for 6 months with follow-up voice calls placed 1, 4, and 6 months after the date of enrollment to assess key study-related outcomes on knowledge transfer and acquisition related to modern contraception, partner communication, and spousal uptake of family planning. The retention rate of the study participants over the 6-month period is shown in Figure 1.

Figure 1. Retention Rates of Study Participants Over the 6 Months Period.



Measures of Study Participant Engagement

A total of 26,988 SMS messages were sent out over the 6-month study period, with an average of 66 messages received by each study participant. Out of the 450 men (the primary study participants) onboarded into the messaging system, 426 (95%) successfully received the messages and only 24 (5%) reported having not received them. The messages were interrupted with periodic quizzes to assess knowledge transfer and acquisition, using a total of 9 questions sent out on a weekly basis. The average response (ie, participation) rate in the weekly quizzes was 23%. We noted an 18.1% increase in couple communication attributed to the mTIP, as the couples opened up to each other more on matters concerning family planning, as highlighted in the sample responses below.

Study Participants' Preferences

Interview findings revealed that the men's telehealth information package was effective and as such, feasible and acceptable for empowerment regarding family planning. As Participant 3 affirmed, "When you read these messages, you find directions...you are guided...the person cares."

The messages were also timely, clear, educational, captivating, and laden with wise counsel, building confidence regarding family planning. They were also shareable, making it possible for the men to share the information. As Participant 3 affirmed, "It has given us confidence about family planning...when you get information from medical personnel...you are in a comfortable position to practice it."

Additionally, the questions within the messages were not only stimulatory, but also enhanced discussion and reflection to deepen understanding. Further, the language of the messages was commended for its simplicity and clarity, accentuating access to family planning. Notwithstanding this, the respondents emphasized the need to have the messages translated into the local language to increase access, especially for those that may not be able to read and write.

The improvement in couple communication stemming from this study made it plausible for the spouses to open up to each other, rather than secretly take up family planning. The couples had candid conversations on family size, spacing, finances, and their family planning options. As Participant 1 explained, "this kind of study has enabled us...[to] come together as partners. You know in our local setting; we don't want to share this kind

of information with our wives...so this...has enabled us [to] realign our education about family planning."

The men were empowered to support their wives regarding family planning, creating harmony in the home. Some men were relieved that their wives were using family planning, which according to Participants 9's narrative was a relief from the potential financial strain of a large family: "I started family planning, now my second born is two years and he [her husband] sees it as very good. He is going to plan for them well, and that space is enough for a child to grow well without falling sick all the time...he is happy."

Discussion

Male Involvement in Family Planning

Innovations addressing male involvement in SRH and family planning services have mainly centered on structural barriers by extending clinic hours, allocating specific clinic times for men, and using male champions, among other strategies [13,14]. However, the need to target men with informational and behavioral messages on family planning by leveraging channels that reach them where they are located (eg, homes, workplaces, or bars) is still new, with digital solutions taking center stage [14,15].

Our study assessed the acceptability and feasibility of an mTIP that leveraged a toll-free hotline and SMS messages as channels to disseminate information on SRH and modern contraception. The 95% success rate for mobile message dissemination (426 of the 450 men successfully received the messages) shows the potential digital platforms have as effective channels for cascading family planning information to target audiences. This is especially important in reaching men who are often left out from traditional physical and mass media campaigns, as these operate in spaces that are largely seen as spaces for women, require lengthy contact time, and are not customized to meet the individual needs of men [16,17].

The 6-month retention rate of study participants in the virtual cohort was 399/450 (88.7%), positioning digital channels like SMS and voice calls as effective and sustainable platforms for continuous engagement beyond physical locations. This is supported by the rising number of people in Uganda who own mobile phones; that number stood at approximately 26 million in December 2020 [18]. In addition, the virtual cohort offered an opportunity for follow up beyond the confines of the health facility or community outreach activities, which are the traditional entry points to accessing family planning services in Uganda.

The relatively high mobile phone ownership rate among the spouses (65/101, 63.9%) offered an opportunity to diversify the digital innovations that can be deployed in the space of family planning. For example, gamified mobile applications that assess decision-making skills and knowledge transfer have been noted to stimulate more engagement with users [19,20]. The 23% participation rate of study participants in the weekly quizzes was relatively low even when compared to other studies that leveraged SMS quizzes during end user assessments. For example, the "Text-to-Change" study, which used SMS messages to disseminate information targeted at youth on HIV and AIDs, had a 53% average participation rate [21]. The discrepancy might be attributable to the absence in our study of complementary media platforms, such as flyers or a radio campaign, to increase awareness and boost participation.

Using simple, rather than complex, language or terminology in developing mobile health messages is important for end users to easily interpret the messages. This was revealed through interviews with the participants, who commended the simple, comprehensible family planning messages. In a setting where health care delivery models leave men out, owing to their work schedules and negative health-seeking behaviors, innovations that engage them within their comfort zones will be instrumental in overcoming barriers to health care access embedded within patriarchal societies in sub-Saharan Africa. This is especially important in our context, where women have traditionally sought for permission and support from their partners, in the form of transport and time, in order to access SRH services such as family planning [22,23].

In addition, given the limited contact time and space and the inadequate customization of traditional media and interpersonal communication models to suit specific local demographics for health information, current trends in mobile phone ownership in Uganda [20] offer the opportunity to leverage these ubiquitous tools for health information dissemination and reach larger audiences with minimal investment.

Principal Results

Out of the 450 men (the primary study participants) onboarded onto the family planning mobile message plan, 426 (95%) successfully received the messages and only 24 reported having not received them. The average response (ie, participation) rate in the weekly quizzes was 23%. There was a noted 18.1% increase in couple communication attributed to the mTIP, and the couples opened up to each other more on matters concerning family planning.

Limitations

Periodic outages of the SMS system inhibited the receipt of some of the scheduled family planning messages, disrupting information access flow. This was addressed by setting up an alert system for outages that enabled the software developers to be notified early enough for quick resolution with minimal disruption. Additionally, the unavailability of some of the study participants' phones during the scheduled phone calls disrupted communication. This was addressed by rescheduling the follow-up call on an alternative day within the follow-up window. The study achieved 95% (551 of 576) of its target sample size, which was slightly deficient, but only negligibly affected the statistical power of the results. Notwithstanding this, the study will serve as a pilot study for a future large, randomized controlled trial of mobile phones as a channel for disseminating information on family planning, to truly measure the impact of digital telehealth as a channel for promoting family planning.

Comparison With Prior Work

The use of mobile phones as a channel for disseminating information to bring about behavioral change has gained momentum in sub-Saharan Africa. The use of hotlines, SMS messages, mobile phone apps, and social media have been documented. In most cases, one or a combination of these channels is used to cascade health information to the target audiences, with the desired behavioral change outcomes being tracked. SMS messages have been extensively deployed in different public health programs for patient education and self-awareness of noncommunicable diseases, reminder systems in maternal health to promote antenatal attendance, and by health systems to strengthen the performance of health workers [22,23].

Therefore, the choice of SMS messages in our study as a channel for engaging with the study participants was closely informed by similar past programs. As a measure to curb message fatigue among recipients, past SMS interventions have limited the number of messages sent out to an average of 1 to 2 per week, similar to the approach taken in our study. A review of demographic health data on SMS-based family planning communication within low- and middle-income countries showed an uptake of about 5.4% within selected African countries [24]. This low utilization and uptake mirrors our 24% average participation in the periodic quizzes. Methods for assessing interventional impact in most studies have involved administering before and after interviews. For our study, we opted to perform interviews during the study at 1, 4, and 6 months from the enrollment date, in order to track changes in the outcome indicators. We believe this helped to rule out any possible impact from confounding factors that could have arisen from one-time assessment surveys or interviews.

Conclusions

Digitally supported communications channels (SMS messages and phone hotlines) for disseminating health information on family planning could be leveraged for a wider reach with minimal resource input given limited contact time and space and the capacity for customization of the message to specific demographics. Digitally supported communication channels can provide ways to address participants' concerns and inquiries in real time, or as fast as possible, increasing the likelihood of adoption of family planning among couples. There is a need for additional studies on the influence of mobile messaging on behavioral changes.

Acknowledgments

This study was funded by The National Academies of Sciences, Engineering, and Medicine under the Partnership for Enhanced Engagement in Research funding mechanism. The authors appreciate the support provided by the doctors at The Medical Concierge Group call center, who handled the participants' inquiries.

Authors' Contributions

JB, JK, and JA led the data collection and cleaning process. VN led the quantitative data analysis, LN led the qualitative data analysis, and LHK led the manuscript preparation, writing, and review process. JMB, KJH, DM, and AK contributed to the study design, manuscript review, and approval of the final manuscript version.

Conflicts of Interest

None declared.

References

1. National population and housing census 2014: main report. Uganda Bureau of Statistics. 2015. URL: https://www.ubos.org/wp-content/uploads/publications/03_20182014_National_Census_Main_Report.pdf [accessed 2022-05-30]
2. Uganda Demographic and Health Survey 2016. The Demographic and Health Surveys Program. 2018. URL: https://dhsprogram.com/pubs/pdf/FR333/FR333.pdf [accessed 2022-05-30]
3. Götmark F, Andersson M. Human fertility in relation to education, economy, religion, contraception, and family planning programs. BMC Public Health 2020 Feb 22;20(1):265 [FREE Full text] [doi: 10.1186/s12889-020-8331-7] [Medline: 32087705]
4. Adserá A. International political economy and future fertility trends. Vienna Yearb Popul Res 2020;18:37-32 [FREE Full text] [doi: 10.1553/populationyearbook2020.deb01]
5. Khan S, Bradley S, Fishel J, Mishra V. Unmet need and the demand for family planning in Uganda: Further analysis of the Uganda Demographic and Health Surveys, 1995-2006. The Demographic and Health Surveys Program. Calverton, MA: Macro International; 2008. URL: http://dhsprogram.com/pubs/pdf/FA60/FA60.pdf [accessed 2022-06-01]
6. Kaneda T, Smith R. Unmet Need for Family Planning: What Can We Learn From the DHS 5-Year Contraceptive Calendar Data? Population Reference Bureau. URL: https://www.prb.org/resources/unmet-need-for-family-planning-what-can-we-learn-from-the-dhs-5-year-contraceptive-calendar-data/ [accessed 2022-05-30]
7. Potasse MA, Yaya S. Understanding perceived access barriers to contraception through an African feminist lens: a qualitative study in Uganda. BMC Public Health 2021 Feb 02;21(1):267 [FREE Full text] [doi: 10.1186/s12889-021-10315-9] [Medline: 33530960]

8. Sileo K, Wanyenze R, Lule H, Kiene. Determinants of family planning service uptake and use of contraceptives among postpartum women in rural Uganda. Int J Public Health 2015 Dec;60(8):987-997 [FREE Full text] [doi: 10.1007/s00038-015-0683-x] [Medline: 25967466]
9. askNivi: Applying behavioral science, design, and technology to address consumer-facing health needs in emerging markets. Digital Health Compendium. URL: https://fpdigitalsolution.org/case-study/asknivi/ [accessed 2022-05-30]
10. Njagi J. Delivering sexual and reproductive health education to girls: Are helplines useful? Girlhood Stud 2018;11(2):30-45. [doi: 10.3167/ghs.2018.110204]
11. van Heerden A, Norris S, Tollman S, Richter L, Rotheram-Borus M. Collecting maternal health information from HIV-positive pregnant women using mobile phone-assisted face-to-face interviews in Southern Africa. J Med Internet Res 2013 Jun 10;15(6):e116 [FREE Full text] [doi: 10.2196/jmir.2207] [Medline: 23748182]
12. L'Engle KL, Vahdat HL, Ndakidemi E, Lasway C, Zan T. Evaluating feasibility, reach and potential impact of a text message family planning information service in Tanzania. Contraception 2013 Feb;87(2):251-256. [doi: 10.1016/j.contraception.2012.07.009] [Medline: 22935322]
13. Rongkavilit C, Naar-King S, Kaljee LM, Panthong A, Koken JA, Bunupuradah T, et al. Applying the information-motivation-behavioral skills model in medication adherence among Thai youth living with HIV: a qualitative study. AIDS Patient Care STDS 2010 Dec;24(12):787-794 [FREE Full text] [doi: 10.1089/apc.2010.0069] [Medline: 21091238]
14. Dubov A, Altice F, Fraenkel L. An information-motivation-behavioral skills model of PrEP uptake. AIDS Behav 2018 Nov;22(11):3603-3616. [doi: 10.1007/s10461-018-2095-4] [Medline: 29557540]
15. Pascoe L, Herstad M, Shand T, van den Heever L. Building Male Involvement in SRHR: A basic model for Male Involvement in Sexual and Reproductive Health and Rights. XYonline. 2012. URL: https://xyonline.net/sites/xyonline.net/files/Pascoe%2C%20Building%20male%20involvement%20in%20SRHR%202012_0.pdf [accessed 2022-05-30]
16. Eqtait FA, Abushaikha L. Male involvement in family planning: an integrative review. Open J Nurs 2019;09(03):294-302 [FREE Full text] [doi: 10.4236/ojn.2019.93028]
17. Ippoliti NB, L'Engle K. Meet us on the phone: mobile phone programs for adolescent sexual and reproductive health in low-to-middle income countries. Reprod Health 2017 Jan 17;14(1):11 [FREE Full text] [doi: 10.1186/s12978-016-0276-z] [Medline: 28095855]
18. Jooste K, Amukugo HJ. Male involvement in reproductive health: a management perspective. J Nurs Manag 2013 Mar;21(2):327-338. [doi: 10.1111/j.1365-2834.2012.01332.x] [Medline: 23409867]
19. Adelekan A, Omoregie P, Edoni E. Male involvement in family planning: challenges and way forward. Int J Popul Res 2014 May 26;2014:1-9 [FREE Full text] [doi: 10.1155/2014/416457]
20. Market Performance Report 4Q2020. Uganda Communications Commission. URL: https://www.ucc.co.ug/wp-content/uploads/2021/04/UCC-Q4-2020-Market-Perfomance-Report-compressed.pdf [accessed 2022-05-30]
21. Lupton D, Thomas GM. Playing pregnancy: the ludification and gamification of expectant motherhood in smartphone apps. M/C J 2015 Oct 01;18(5) [FREE Full text] [doi: 10.5204/mcj.1012]
22. Thummalachetty N, Mathur S, Mullinax M, DeCosta K, Nakyanjo N, Lutalo T, et al. Contraceptive knowledge, perceptions, and concerns among men in Uganda. BMC Public Health 2017 Oct 10;17(1):792 [FREE Full text] [doi: 10.1186/s12889-017-4815-5] [Medline: 29017539]
23. Kriel Y, Milford C, Cordero J, Suleman F, Beksinska M, Steyn P, et al. Male partner influence on family planning and contraceptive use: perspectives from community members and healthcare providers in KwaZulu-Natal, South Africa. Reprod Health 2019 Jun 25;16(1):89 [FREE Full text] [doi: 10.1186/s12978-019-0749-y] [Medline: 31238960]
24. Earle S, Marston HR, Hadley R, Banks D. Use of menstruation and fertility app trackers: a scoping review of the evidence. BMJ Sex Reprod Health 2021 Apr 06;47(2):90-101. [doi: 10.1136/bmjsrh-2019-200488] [Medline: 32253280]

Abbreviations

AI: artificial intelligence
IBM: information-behavior-motivation
mTIP: Men's Telehealth Information Package
SMS: short message service
SRH: sexual and reproductive health
TMCG: The Medical Concierge Group

Edited by A Mavragani; submitted 22.10.21; peer-reviewed by VM Kiberu, M Lotto; comments to author 30.11.21; revised version received 05.01.22; accepted 17.03.22; published 28.06.22.

Please cite as:
Kamulegeya LH, Bwanika J, Banonya J, Atuhaire J, Musinguzi D, Nakate V, Kyenkya J, Namatende L, Horvath KJ, Kiragga A Feasibility and Acceptability of a Ugandan Telehealth Engagement Platform for Informational Messaging on Modern Contraception: Pilot Cross-sectional Study
JMIR Form Res 2022;6(6):e34424
URL: https://formative.jmir.org/2022/6/e34424
doi:10.2196/34424
PMID:35763336

Original Paper

Physical Exercise Program on Fall Prevention Using Technological Interface: Pretest Study

Mª Nilza Nogueira[1*], PhD; Joana Silva[2*], PhD; Isabel Nogueira[1], MSc; Maria Neto Pacheco[1], MSc; Joana Lopes[2*], MA; Fátima Araújo[1*], PhD

[1]Escola Superior de Enfermagem do Porto, Centro de Investigação em Tecnologias e Serviços de Saúde, Porto, Portugal

[2]Center for Assistive Information and Communication Solutions, Fraunhofer Portugal Research, Porto, Portugal

[*]these authors contributed equally

Corresponding Author:
Mª Nilza Nogueira, PhD
Escola Superior de Enfermagem do Porto
Centro de Investigação em Tecnologias e Serviços de Saúde
Escola Superior de Enfermagem do Porto
Porto, 4200-072
Portugal
Phone: 351 225 073 500
Fax: 351 225 096 337
Email: nilza@esenf.pt

Abstract

Background: Prevention of falls among older adults has boosted the development of technological solutions, requiring testing in clinical contexts and robust studies that need prior validation of procedures and data collection tools.

Objective: The objectives of our study were to test the data collection procedure, train the team, and test the usability of the FallSensing Games app by older adults in a community setting.

Methods: This study was conducted as a pretest of a future pilot study. Older adults were recruited in a day care center, and several tests were applied. Physical exercise sessions were held using the interactive FallSensing Games app. Nurse training strategies was completed.

Results: A total of 11 older adults participated. The mean age was 75.08 (SD 3.80) years, mostly female (10/11, 91%) and with low (3-6 years) schooling (10/11, 91%). Clinically, the results show a group of older adults with comorbidities. Cognitive evaluation of the participants through the Mini Mental State Examination showed results with an average score of 25.64 (SD 3.5). Functional capacity assessed using the Lawton Instrumental Activities of Daily Living Scale (overall score from 0-23, with lower scores reflecting worse capacity to perform activities) showed impairment in different instrumental activities of daily living (average score 14.27). The data collection tool proved to enable easy interpretation; however, its structure needed small adjustments to facilitate the data collection process. Despite the length of the questionnaire, its implementation took an average of 21 minutes. For the assessment of the prevalence of fear of falling, the need to add a question was identified. The performance of functional tests under the guidance and presence of rehabilitation nurses ensured the safety of the participants. The interactive games were well accepted by the participants, and the physical exercises allowed data collection on the functionality of the older adults, such as the number of repetitions in the tests, range of movement (angle), duration of the movements, and execution of each cycle. Concerning the training of the nurses, it was crucial that they had experience with the platform, specifically the position of the chair facing the platform, the position of the feet, the posture of participants, and the use of sensors.

Conclusions: In the future pilot study, the researchers point out the need to design a study with mixed methods (quantitative and qualitative), thus enriching the study results.

(JMIR Form Res 2022;6(6):e26196) doi:10.2196/26196

KEYWORDS

functional tests; fall assessment; exercices; older adults: games; technology

Introduction

In community settings, fall episodes are highly prevalent among the population of older adults [1,2]. Regardless of the severity of the related injuries, the impact on health and quality of life of older adults and their families can be significant since they often trigger and accelerate a cycle of restrictions and barriers leading to the dependency of the older person for activities of daily living [3-5]. Evidence suggests that the frequency of falls increases with age and degree of fragility [6] and that the presence of risk factors directly influences the risk of fall [7,8].

A structured and standardized screening and assessment of the risk of fall in older adults contribute to its prevention and reduction and are central to the design of the intervention and risk monitoring [7].

Changes in gait and balance are factors that have been strongly associated with the outcome of fall in older adult population, and rapid tests, such as the 30-Second Chair Stand Test (30CST) [9], the 4-Stage Balance Test (4SBT) [10,11] and the Timed Up and Go (TUG) test [7,12], are recommended for their assessment.

Evidence-based fall prevention programs have demonstrated a significant reduction in fall risk, falls, and related injuries in older people in a community setting [13,14]. An exercise program with proven effectiveness in preventing falls is the Otago Exercise Program (OEP), designed at the University of Otago Medical School [15-21]. The focus of the OEP is to improve strength and balance with a simple, affordable, and easy home-implemented solution for 12 months, monitored by a health professional through monthly telephone interviews and biannual home visits.

Recent evidence has strived to integrate technologies into physical exercise programs that have shown a positive effect in adherence and overcoming barriers to exercise, as well as improvements in physical functioning [22,23]. Some technological solutions to facilitate the process of monitoring and fall prevention have already been developed in Portugal, such as the FallSensing Screening and FallSensing Games apps, designed by Fraunhofer Center for Assistive Information and Communication Solutions (AICOS) Portugal.

The FallSensing Screening app uses inertial measurement units (IMUs) to extract information about the user's movements, using these data to characterize the movement; it then uses metrics calculated after processing the sensor signal, obtained during the execution of the functional tests performed by the user. The IMU, composed by a triaxial accelerometer, triaxial gyroscope, and triaxial magnetometer, was used to acquire inertial data during the exercises at 50 Hz. Data were transmitted using Bluetooth Low Energy wireless technology to a main computer where the processing occurs. The interactive FallSensing Games app, based on the OEP, aims to improve physical functionality and is also used as a motivator for participants who perform the exercises.

Therefore, there is an excellent opportunity and a need to develop new user-tailored solutions supported on more robust and valid fall risk predictive models and good clinical practice in fall prevention [6]. Technological solutions need validation in a clinical context, through methodologically robust experimental studies.

In research, pretesting is an essential stage before the pilot study because it allows for identifying weaknesses in the development of measurement instruments (structure, content, semantics) to determine the potential respondents' difficulty in interpreting the questions and complexity of the evaluation process. In addition, it enables benchmarking and training procedures and standardizes the modus operandi in data collection, thus contributing to improving the reproducibility and accuracy of measurements [24,25]. This study aims to test the data collection procedure, train the team, and test the usability of the FallSensing Games app by older adults in a community setting.

Methods

Study Design

A pretest study was performed for the successful implementation of a future larger pilot study. Two research centers were involved in the project, the Nursing School of Porto (ESEP) and the Fraunhofer AICOS Portugal.

Participants

Participants were recruited in one of the day centers in western Porto city. For the realization of the FallSensing Games, a minimum of 6 participants was required, but in this study the sample included 11 older adults.

The inclusion criteria were being aged 65 years or older, living at home, walking independently, not presenting with cognitive impairment according to the Portuguese version of the Mini Mental State Examination (MMSE) [26], not having severe visual or hearing impairment, signing informed consent, and presenting moderate to high risk of falling (assessed through 4 short questions with a dichotomous answer option).

Participant exclusion criteria included having chronic or acute illness conditions for which exercise is contraindicated; ever having hip or knee surgery or having a history of lower limb fractures in the last 12 months; having participated in physical exercise programs in the last 12 months; having participated in another research study; or having a final MMSE score below 22 (with up to 2 years of school), below 24 (3 to 6 years of school), or below 27 (7 years or more of school).

Materials

After selection criteria application, data were collected by the main researchers, and functional testing was performed by two rehabilitation nurses who had received training sessions. In accordance with the best practices recommended for clinical research, the main researcher ensures that their team is trained to implement the different procedures at the different stages of the investigation process. Thus, the training of rehabilitation nurses was performed by the principal investigators. A meeting was held with all the investigators and rehabilitation nurses to present the investigation plan, provide the study dossier (research project, assessment instruments for functional testing instruments, and Otago exercise manual), introduce technologies, explain data collection procedures, and review

communication techniques. In the end, there was time for clarification. Subsequently, a training session was held (Figure 1).

The training procedures took place for a week, covering use of the data collection instrument and performance of functional tests and physical exercises sessions. Among investigators, after consensus meetings, 2 researchers used the data collection instrument, independently and randomly among the participants, in similar spaces. The monitoring of the instrument application time and use of field notes to document difficulties and other observations, such as opinions made by the respondents, were the resources used to assess the applicability of the data collection instrument. Before moving on to the physical exercise sessions, the researchers met to analyze and decide by consensus the questionnaire items identified as needing improvement. Participants performed the functional tests on 3 consecutive days to avoid interfering with the activities previously planned by the day center (Figure 2).

Figure 1. Training procedures.

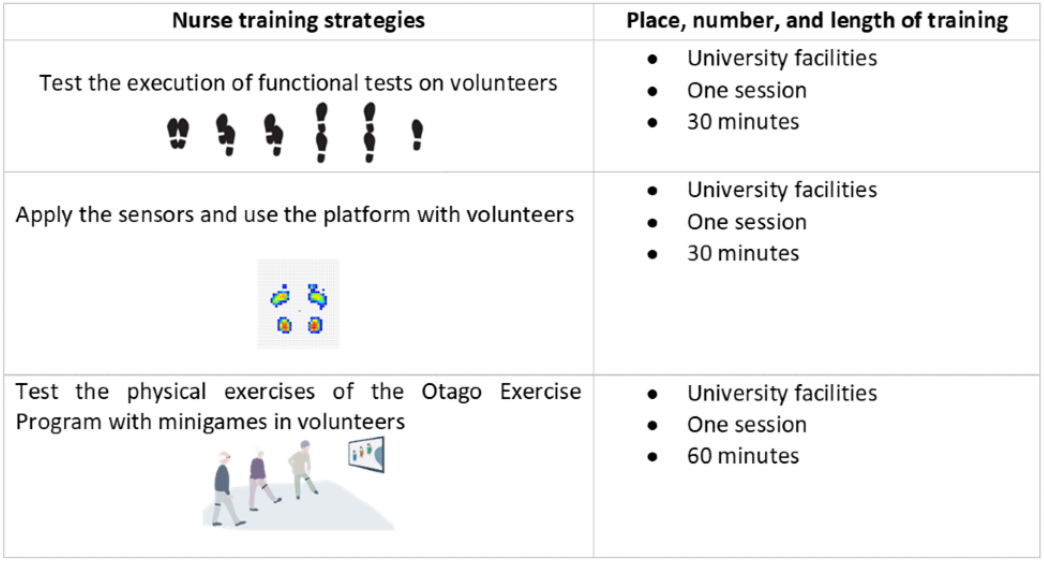
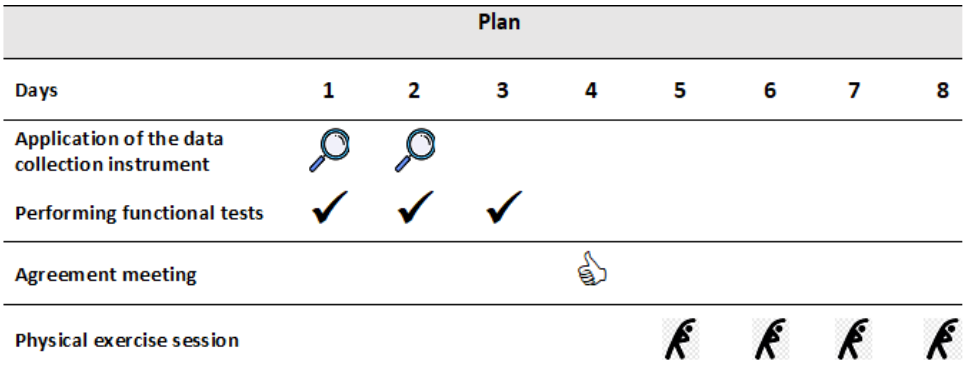


Figure 2. Study plan.



Instrument

The data collection instrument included sociodemographic, clinical, and functional evaluation; fear of falling, and the acceptance of technology. Lower limb strength and muscle resistance were the functional variables assessed through the 30CST; mobility was evaluated by the TUG test (normal step). These functional tests and the 4SBT allowed evaluation of the risk of fall, which was also assessed with the Fall Risk Screening Tool. The functional capacity for activities of daily living was assessed using the Lawton Instrumental Activities of Daily Living Scale (IADL), fear of falling using the Falls Efficacy Scale–International (FES-I), and acceptance of technology by participants and health professionals by the System Usability Scale (SUS). The domains of the instrument are presented below.

30CST Instrument

The performance in the 30CST is used as a measure of the strength and muscle resistance of the lower limbs, specifically the extensor muscles of the knee [27,28]. It is a quick test without a dynamometer, training, or special equipment, which allows evaluating the strength of the lower limbs by counting the number of times the individual stands and sits in 30 seconds [9-30]. More strength in the lower limbs is associated with better balance [9-30], and the functional improvement of older adults after a fall prevention program will be manifested by a greater number of repetitions in 30 seconds in the posttest assessment [14]. The results have shown good psychometric qualities [9-30].

TUG Test (Normal Step)

Since mobility assessment of older adults is a central component in the geriatric assessment [31], the TUG test is proposed to evaluate the clinical utility of the timed stand and walk test. This test measures in seconds the time an individual takes to stand from a chair, walk a distance of 3 meters, return, and sit back in the chair. These authors reported that time spent on the TUG test performance was related to scores on the Berg Balance Scale (r=–0.72) and the walking speed (r=–0.55) and Barthel Activities of Daily Living Index (r=–0.51) scores. Individuals who completed the test in less than 20 seconds were independent in transferring, and individuals who completed the test in more than 30 seconds tended to be dependent on this task.

The TUG test has been widely referred to and used [12] as a screening test to assess the risk of fall in older adults in community settings, namely through the guideline of the American Geriatric Society and British Geriatric Society and in the US Centers for Disease Control and Prevention (Stopping Elderly Accidents, Deaths, and Injuries [STEADI initiative]).

A prospective design conducted to evaluate the predictive ability of the TUG test for future falls and estimate the best cutoff point of the test pointed to 12.6 seconds, with the corresponding values of sensitivity (30.5%), specificity (89.5%), positive predictive value (46.2%), and negative predictive value (81.4%) [12]. The researchers who conducted the study emphasized the high specificity (89.5%) and high negative predictive value (81.4%) to a cutoff point of 12.6 seconds as a support for the clinical utility of this test in older adults at high risk of falling. Researchers reported that after a fall prevention program, performing the test in less time is indicative of improvement [14]. In Portugal, this test has been used in several studies [32,33].

4SBT Instrument

The balance tests were conceptually developed to track balance impairments [11,34,35] placing the older adults at risk of falling [10,11]. More specifically, the 4SBT is used to track impairment in the static balance of older adults. Several authors have found the test with an excellent performance in test-retest reliability (r=.97) and interrater reliability (K=.92) [10,36]. The success of fall prevention programs is measured by comparing the positions achieved in 10 seconds in the pre- and postprogram evaluation [14]. The final score will be the number of positions successfully completed for 10 seconds without losing balance. The older adults who cannot maintain position 3 for 10 seconds have a high risk of falling [37].

IADL Instrument

The IADL assesses the level of independence of older adults in performing activities of daily living, which integrate day-to-day tasks such as using the telephone, shopping, preparing food, housekeeping, washing clothes, using transport, preparing medication, and handle finances. It is an easy-to-administer tool that can be used with older adults in a community and hospital setting [38-40].

In this study, the Portuguese version [41], which uses the same items as the original version but applies a polychotomous score (0, 1, 2, 3, or 4) instead of the original dichotomous score (0 and 1), was used, allowing for a better description of a person's ability to perform the tasks, giving each response option a different score. The total score of the scale varies from 0 to 23, with a lower score corresponding to worse performance. In the validation study, the instrument showed good metric qualities to be applied in a community setting.

FES-I Instrument

The fear of falling among older adults is an expressive problem and highly relevant because it is associated with adverse effects on mobility and quality of life [42-44]. One of the instruments used to evaluate this construct is the FES-I [45]. Its adaptation to different languages and cultural contexts (following the protocol recommended by the Prevention of Falls Network Europe Group), has allowed the instrument to be widely applied and the results compared in different populations and contexts. The FES-I version is an instrument that incorporates some daily activities that are a little more complex than those of the original version and others more focused on the social life of older adults as a way to overcome some weaknesses identified in the original version. For each of the 16 items, the answer option is based on a 4-point Likert scale (1=not at all worried; 2=somewhat worried; 3=moderately worried; 4=very worried). The instrument has shown validity, reliability, and comparability across cultures, so it is recommended for research practice and the clinical context, namely in fall prevention programs for older adults population [46].

In this research, the FES-I version validated for the Portuguese population [47] showed excellent internal consistency (α=.98) and test-retest reliability (intraclass correlation coefficient 2.1=0.999). Concurrent validity, assessed using the Activity-Specific Balance Confidence Scale, presented results indicative of good concurrent validity (r_s=–0.85; P<.001). Considering the global results, the authors consider the Portuguese version of the FES-I a reliable and valid measure to assess the fear of falling among the Portuguese older adult population living in the community.

SUS Instrument

To validate the acceptance of the technology by participants and health professionals, the responses of the SUS was analyzed. This rapid test evaluates the usability of a certain product or service [48]. This test has several features that provide a good assessment of the overall usability, such as the flexibility to evaluate interface technologies, interactive voice response systems [49], hardware platforms used in more traditional computer interfaces, and websites. Ease and speed of use (by both participants and system administrators), ease of operation of scoring, and the free access characteristic are also advantages.

The original SUS instrument consists of 10 statements that are scored on a 5-point Likert scale (1=strongly disagree to 5=strongly agree) [48]. The final score can vary from 0 to 100 points, with a better score indicating better usability [49], and the final score needs to be considered following the instructions of the original instrument because statements switch between positive and negative. A study conducted at the national level performed a psychometric analysis of the tool intending to

translate, culturally adapt, and contribute to the validation of the European Portuguese version of the SUS [50].

Procedures

Technology Platform

The technology platform uses mobile app inertial sensors to extract information on movement performed by the participant and related characterization. This platform measures pressure distribution at 50 Hz and comprises 1600 pressure sensors (10 mm × 10 mm) with maximum value of 100 N/sensor. The size of the active area of the pressure platform is a square matrix of 40 cm × 40 cm. Voltage data are converted with an 8-bit A/D converter and transmitted via USB to the main computer. The risk of fall is then determined from parameters calculated after processing the signal from the inertial sensors during the execution of functional tests such as walking, sitting, and standing.

Games

The interactive FallSensing Games app is based on the OEP and the use of inertial sensors to monitor the movements performed by the participants during these exercises. To interact with the characters and achieve the objectives of each game, participants must perform the suggested movements correctly. Monitoring the movements of participants in each game allows us to assess the evolution of physical capacity and extract parameters related to functional capacity.

The FallSensing Games app comprises 3 minigames, with each minigame comprising 2 to 3 exercises from the OEP. The composition of the minigames is as follows [29]:

- Minigame 1 includes knee bends and sit-to-stand exercises monitored with a sensor on the thigh.
- Minigame 2 includes lateral hip abduction (side hip), frontal knee extension (front knee), and backwards knee flexion (back knee) monitored with the sensor on the ankle.
- Minigame 3 includes calf and toe raises monitored with sensors on the top of the foot.

Physical Exercise Session

The Otago physical exercise session, supported by interactive games, was implemented by the rehabilitation nurses in the day centers. Participants were divided into 2 groups. In the physical exercise session, after the demonstration, the rehabilitation nurse started with the warm-up exercises suggested by the OEP followed by exercises to strengthen the lower limbs and improve balance and stability and finally the relaxation phase, with stretching exercises. For the implementation of the physical exercise session, we used the following material resources: (1) room with free space (at health centers), (2) computer and television/projector, (3) wearables (IMUs) with loaders and fixing tapes, and (4) a pressure platform.

Ethics Approval

The pretest study was approved by the Health Ethical Committee from ESEP (annex 5 to document no. 4/2019). All participants were informed and provided informed consent in duplicate (one copy for participant and one copy to investigator) before enrolling. Participants were informed about the confidential information protection, the right to study withdrawal, data anonymity, and the likelihood of study publication.

Data Analysis

SPSS (version 26.0, IBM Corp) software was used for statistical analysis. The univariate descriptive analysis was applied to describe data supported on measures of central tendency and dispersion.

Results

Descriptive Information

The pretest was conducted on 11 participants with a mean age of 75.08 (SD 3.80) years, mostly female (10/11, 91%) and with low (3-6 years) schooling (10/11, 91%). Clinically, the results show a group of older adults with comorbidities who portray the epidemiological profile of chronic disease, with high expressiveness of hypertension, osteoarticular disease, and urinary incontinence. In this sample, despite the high prevalence of osteoarticular disease, only 2 older adults used walking aids. More than half (6/11, 55%) of participants reported depression. This clinical pattern was accompanied by drug regimens that integrate mostly 4 or more drugs (10/11, 91%). Balance impairment or difficulty in walking was referred to by 64% (7/11) of participants. More than half (6/11, 55%) presented a high risk of falling, due to recurrent falls in the last 12 months.

Cognitive evaluation of the participants using the MMSE showed results with an average score of 25.64 (SD 3.5), consistent with a mild degree of impairment for participants with a low level of education. Functional capacity assessed using the IADL (overall score varying from 0 to 23, with lower scores reflecting worse capacity to perform activities) showed impairment in different activities of daily living (average score 14.27). The results of the descriptive analysis for sociodemographic and clinical variables are presented in Table 1.

Concerning the fear of falling, the results showed that the activities in which participants reported higher levels of fear (response options: 3=moderately concerned and 4=very concerned) were walking on slippery surfaces (7/11), going up and down stairs (6/11), and walking on uneven surfaces and walking up and down slopes (5/11 for both). In the self-care dressing/undressing, shopping, and walking in the neighborhood, 4 older adults were identified with the response option 3 or 4 on the Likert scale. The response option equal to 2 (a little worried) was expressed by 10 out of 11 participants for a variable number between 1 to 6 activities.

Table 1. Participant characteristics.

Characteristics	Value
Age (years), mean (SD)	75.09 (3.80)
65-74, n (%)	3 (27)
75-84, n (%)	8 (73)
Gender (female), n (%)	10 (91)
Marital status, n (%)	
Married	2 (18)
Single	2 (18)
Divorced	2 (18)
Widowed	5 (45)
Education (years), n (%)	
0-2	1 (9)
3-6	10 (91)
Comorbid health conditions (yes), n (%)	
Arterial hypertension	8 (73)
Osteoarticular disease	8 (73)
Urinary incontinence	8 (73)
Depression	6 (55)
Vertigo syndrome	6 (55)
Diabetes mellitus	5 (45)
Vision changes	4 (36)
Daily medication consumption (≥4), n (%)	10 (91)
MMSE[a] score, mean (SD)	25.64 (3.50)
IADL[b], mean (SD)	14.27 (7.14)
Walking difficulties/balance compromised (yes), n (%)	7 (64)
Walking aids (yes), n (%)	2 (18)
Falls (yes), n (%)	
High risk	6 (55)
History of falls (last 12 months)	6 (55)
Recurrent falls	6 (55)
Indoor falls	4 (67)
Health care need after falls	2 (33)

[a]MMSE: Mini Mental State Examination

[b]IADL: Lawton Instrumental Activities of Daily Living Scale

Data Collection Procedure

In general, at the interview stage no difficulties of interpretation were identified that could make it difficult to answer the questions of the data collection instrument; however, for the FES-I assessment both researchers needed to frequently recall the Likert scale in use. In the consensus meetings, it was found in the field notes of the researchers that the behavior of the older adults in the assessment of fear of falling are indicative of increased difficulties in interpreting the request and choosing the answer option. This fact showed the need to evaluate this construct in a simpler way that also allowed distinguishing the participants regarding the level of fear of falling. After research, it was decided to include a single-item question “Are you afraid of falling?” with the same ordinal answer option (not at all worried, a little worried, moderately worried, very worried) in the definitive questionnaire for the future pilot study. According to some authors, there is not enough evidence that more complex measures consisting of several items better assess the fear of falling into this population range than single item questions [51-53].

In the 30CST test, an average of 4.9 repetitions was obtained. The average time of the TUG test was 21.9 seconds, with 12.6 seconds being used as a cutoff point, less time in the test performance means better functional condition. Finally, in the performance of the 4SBT, all participants were able to perform positions 1 and 2 with their eyes open, but only 55% (xx/xx) were able to achieve position 3 (Table 2).

During the functional tests, the participants presented difficulties in the execution of the instruction given for the 4SBT test, despite the previous demonstration of the 4 foot positions performed by both rehabilitation nurses under the supervision of the investigators (Figure 3).

During the 4SBT functional test, the older adults showed a behavior of constantly searching for support in the surrounding environment (people, walls, chairs). This fact is reported in the literature on falls in elderly populations as indicative of fear of falling or low perception of self-efficacy to perform the task.

The average time to complete the data collection tool—sociodemographic data, clinical and drug consumption history, IADL, and FES-I—was 21 (SD 2.62) minutes. Before starting data collection, each of the researchers reminded the participants of the study objectives and the possibility of being able at any time to express their willingness to withdraw without any negative implication.

Table 2. Functional test results.

Tests	Value	Minimum-maximum
30CST[a], mean (SD)	4.9 (3.315)	0-11
TUG[b]		
Duration (seconds), mean (SD)	21.90 (5.74)	13.96-31.38
≥12.6 seconds, n (%)	100	—[c]
4SBT[d] (4 foot positions), n (%)		
Position 1	100	—
Position 2	100	—
Position 3	55	—
Position 4	9	—

[a]30CST: 30-Second Chair Stand Test.

[b]TUG: Timed Up and Go.

[c]Not applicable.

[d]4SBT: 4-Stage Balance Test.

Figure 3. Number of foot positions in 4-Stage Balance Test, with eyes open.

Train the Team

The training of nurses was meant to standardize the application of functional tests, guidance in games and physical exercises, and interactions with the older adults through appropriate communication techniques and security measures. Regarding the use of technology, the training of nurses allowed validation of the correct use and placement of sensors as well as the correct use of the platform.

Functional Tests Application Procedure

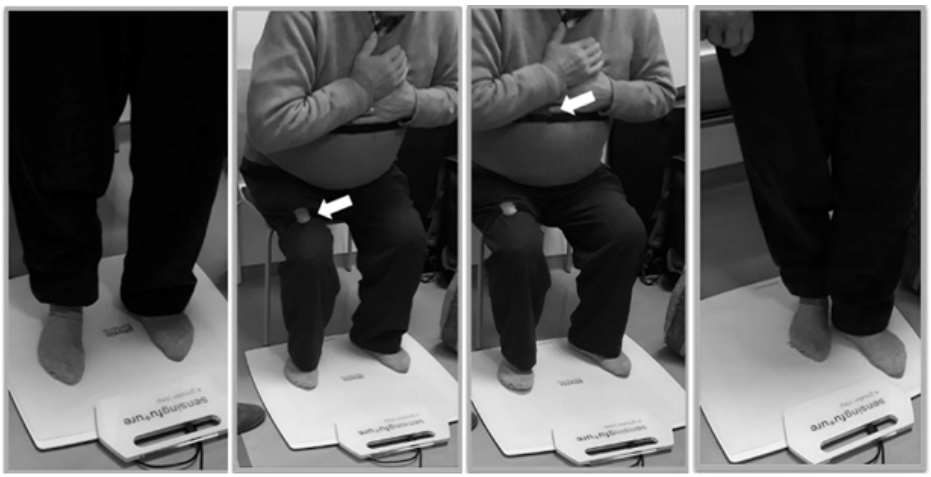
Before the functional tests application procedure, there was a need to establish a relationship of trust between the nurses and participants. The tests were explained and then demonstrated. Special attention was given to the nurses' position toward older adults, tone of voice, rhythm of explanation, and nonverbal communication.

In particular, for the 4SBT it was important to measure the position of the chair facing the platform, the position of the feet, and the posture of older adults on the platform. For both tests (4SBT and TUG), the position of the nurse beside older adults during the test execution proved to ensure the safety of the participants. To mark the path of the execution of the TUG test, colored ribbons were placed on the floor, which were identified by the participants as providing good assistance (Figure 4).

Regarding the use of technological devices, it was necessary to check the position of the sensors, both in the anatomical location and in their local adjustment (avoiding discomfort for older adults or coming loose in order to obtain correct readings and avoid repeating the test several times (Figure 5).

Figure 4. Timed Up and Go test execution path, marked with colored ribbons.

Figure 5. Anatomical position of sensors and foot positions inside the platform.



Physical Exercise Session

The physical exercise session using the FallSensing Games app led to adjustments in its implementation: (1) maintenance of a minimum distance between the participants' chairs to avoid touching each other during the shoulder abduction exercise, (2) synchronization between the material resources (TV/computer) and the start of the warm-up exercises, (3) presence of the nurse near the participants to adjust the correct use of the elastic bands, and (4) adaptation of the nurses' paralanguage (tone of voice) to the sound volume of the games.

Minigames

For each exercise composing 1 of the 3 minigames, the wearable inertial sensors allow us to extract 3 relevant metrics, such as the range of motion angle (angle), range of motion duration (cycle time), and number of repetitions (nr_cycles). Considering that each exercise requires a specific number of repetitions defined according to the OEP, each of these metrics will be computed for each repetition. For example, if we consider the knee bending exercise, each person should perform 10 repetitions of the exercise, allowing us to compute the angle and duration of each repetition and also count the number of performed repetitions. Given that each participant performed each 1 of the 3 minigames, we have computed the mean of each metric for each participant. Table 3 presents the values for each metric averaged for all the participants and its standard deviation.

For minigame 1, each participant performed on average 6 repetitions of sit-to-stand and 23 repetitions of the knee-bending exercise. For minigame 2, each participant performed on average 25 repetitions of knee extension, 25 repetitions of knee flexion, and only 13 repetitions of hip abduction (side hip) exercise. For minigame 3, each participant performed on average 15 repetitions of calf raises and 13 repetitions of toe raises.

In sum, each participant performed on average more repetitions of each exercise than suggested in the OEP due to the gamification of these exercises in the minigames, which

requested the participants to perform more repetitions to accomplish a higher game score. This can be seen as a positive effect of the gamification of the Otago exercises. Another relevant outcome is the retrieval of range of motion–related metrics, as the angle and duration of movements, which can only be quantified when using wearable sensors as opposed to traditional observational programs.

Table 3. Minigame metrics.

Game number and inertial sensor metric	Value, mean (SD)
Minigame 1	
Sit_to_stand_angle	64.43 (22.29)
Sit_to_stand_cycle_time	3.20 (1.46)
Sit_to_stand_nr_cycles	6.33 (4.54)
Knee_bends_angle	35.67 (14.61)
Knee_bends_cycle_time	2.19 (0.82)
Knee_bends_nr_cycles	22.58 (14.44)
Minigame 2	
Knee_extension_angle	80.02 (21.80)
Knee_extension_cycle_time	2.02 (1.37)
Knee_extension_nr_cycles	25.00 (10.39)
Knee_flexion_angle	80.98 (17.84)
Knee_flexion_cycle_time	1.90 (0.98)
Knee_flexion_nr_cycles	24.56 (10.74)
Side_hip_angle	50.03 (26.11)
Side_hip_cycle_time	2.45 (1.11)
Side_hip_nr_cycles	13.00 (5.02)
Minigame 3	
Calf_raises_angle	35.72 (22.92)
Calf_raises_cycle_time	3.46 (4.72)
Calf_raises_nr_cycles	14.75 (9.85)
Toe_raises_angle	29.48 (24.21)
Toe_raises_cycle_time	2.64 (1.86)
Toe_raises_nr_cycles	13.42 (5.98)

Test the Usability of the FallSensing Games App

As for participant satisfaction using the SUS, the results showed that out of the participants who responded, 50% (5/10) assessed the usability of the technology as acceptable, 30% (3/10) expressed good satisfaction, and 20% (2/10) considered the usability of the technology as problematic. One participant did not respond.

In addition to this quantitative analysis, field notes were collected on satisfaction expressed by the participants, who voiced their satisfaction with the games, participation in team games, and repetition of the activity's animated penguins. Statements from participants included "I had never made a game looking at penguins," "I didn't realize that time was passing," "I even forgot the pains," "I even laughed a little bit," and "You need to come here more often." During the stay at the day center, we observed the acceptance of the games, both for the ease of integration in the activities of older adults and for the ease with which older adult followed the games.

Discussion

Principal Findings

Test the Data Collection Procedure

From the nurses' perspective, due to the speed and consistency of the participant answers, the data collection tool proved to enable an easy interpretation. However, its structure needed small adjustments to facilitate the data collection process. Despite the length of the questionnaire, its implementation took an average of 21 minutes. For the assessment of the fear of falling, the need to add a question was identified to clarify whether the participant was afraid of falling. The performance of functional tests by the participants under the guidance and presence of rehabilitation nurses ensured the safety of the participants.

Train the Team

Concerning the training of the nurses, it was crucial that they had experience with the platform, specifically the position of the chair facing the platform, the position of the feet, and the posture of the participants on the platform, which allowed adjustments to minimize errors in the functional test assessment. At the same time, the use of sensors and their anatomical position and adjustment allowed us to understand that the way to hold them needs to be improved.

Test the Usability of the FallSensing Games App

Regarding the games, we can point out 2 aspects. On the one hand, each participant performed on average more repetitions of each exercise than suggested in the OEP to achieve a higher game score. On the other hand, obtaining metrics related to range of motion, such as the angle and duration of movements, was only possible with the use of wearable sensors. The easy integration of games in the activities of the older adult care center and the ease of the older adults in following the games corroborates the results presented by previous research [29].

Limitations

As limitations of the study, we highlight the (1) small sample size; (2) absence of an observation grid of the participants' behavior during the performance of the functional tests and games, which could, in a future pilot study, reflect the realism of the situation under study; (3) spontaneous appreciation of the participants, expressed by the contentment and appreciation of the moments spent together, showing the researchers the need to collect this experience in a planned and rigorous way, namely the feelings and emotions of the participants; and (4) concern to prepare the team of nurses for the application of functional tests, use of the platform and sensors, and physical exercise session with the games led to some aspects being neglected, namely the possibility of incorporating qualitative component into the study.

Therefore, in the future pilot study, the researchers point out the need to design a study with mixed methods (quantitative and qualitative), thus enriching the study results. The researchers, intend to use qualitative methods, such as focus group, for the participants, which can enrich the exchange of experiences during the games and nonparticipant observation, with the use of an observation grid, which can favor the collection of information on the correct execution of the Otago exercises.

Regarding the sample size, the recruitment can be improved by incorporating more day centers and a longer period for data collection.

Despite the limitations of the pretest study and results, this study aims to contribute to the practice of professionals in clinical and research contexts, given the scarcity of information on this relevant stage in experimental/quasi-experimental studies.

Acknowledgments

The National Network of Knowledge Transfer on Medical Devices is an initiative cofunded by Programa Operacional Competitividade e Internacionalização, Portugal 2020, and the European Union through the Fundo Europeu de Desenvolvimento Regional.

Authors' Contributions

NN and FA contributed to the design of the study protocol. NN and FA contributed to the drafting the manuscript. NN, FA, and JS contributed with critical revisions to the paper for important intellectual content. NN, FA and JS obtained the funding. JS described the technological solutions used in the study. NN, FA, IN and MP contributed to the definitions of participant recruitment and ethical considerations.

Conflicts of Interest

None declared.

References

1. Brito TA, Coqueiro RDS, Fernandes MH, de Jesus CS. Determinants of falls in community-dwelling elderly: hierarchical analysis. Public Health Nurs 2014;31(4):290-297. [doi: 10.1111/phn.12126] [Medline: 24862435]
2. Stevens JA, Mahoney JE, Ehrenreich H. Circumstances and outcomes of falls among high risk community-dwelling older adults. Inj Epidemiol 2014 Dec;1(1):5 [FREE Full text] [doi: 10.1186/2197-1714-1-5] [Medline: 27747670]
3. Schoene D, Heller C, Aung YN, Sieber CC, Kemmler W, Freiberger E. A systematic review on the influence of fear of falling on quality of life in older people: is there a role for falls? Clin Interv Aging 2019;14:701-719 [FREE Full text] [doi: 10.2147/CIA.S197857] [Medline: 31190764]
4. Thiem U, Klaaßen-Mielke R, Trampisch U, Moschny A, Pientka L, Hinrichs T. Falls and EQ-5D rated quality of life in community-dwelling seniors with concurrent chronic diseases: a cross-sectional study. Health Qual Life Outcomes 2014 Jan 08;12:2 [FREE Full text] [doi: 10.1186/1477-7525-12-2] [Medline: 24400663]
5. Vitorino LM, Teixeira CAB, Boas ELV, Pereira RL, Santos NOD, Rozendo CA. Fear of falling in older adults living at home: associated factors. Rev Esc Enferm USP 2017 Apr 10;51:e03215 [FREE Full text] [doi: 10.1590/S1980-220X2016223703215] [Medline: 28403369]

6. WHO Global Report on Falls Prevention in Older Age. Geneva: World Health Organization; 2007. URL: https://extranet.who.int/agefriendlyworld/wp-content/uploads/2014/06/WHo-Global-report-on-falls-prevention-in-older-age.pdf [accessed 2022-06-01]
7. Panel on Prevention of Falls in Older Persons, American Geriatrics Society, and British Geriatrics Society. Summary of the updated American Geriatrics Society/British Geriatrics Society clinical practice guideline for prevention of falls in older persons. J Am Geriatr Soc 2011 Jan;59(1):148-157. [doi: 10.1111/j.1532-5415.2010.03234.x] [Medline: 21226685]
8. Kvelde T, Lord SR, Close JCT, Reppermund S, Kochan NA, Sachdev P, et al. Depressive symptoms increase fall risk in older people, independent of antidepressant use, and reduced executive and physical functioning. Arch Gerontol Geriatr 2015;60(1):190-195. [doi: 10.1016/j.archger.2014.09.003] [Medline: 25262556]
9. Cho KH, Bok SK, Kim Y, Hwang SL. Effect of lower limb strength on falls and balance of the elderly. Ann Rehabil Med 2012 Jun;36(3):386-393 [FREE Full text] [doi: 10.5535/arm.2012.36.3.386] [Medline: 22837975]
10. Murphy MA, Olson SL, Protas EJ, Overby AR. Screening for falls in community-dwelling elderly. J Aging Phys Activity 2003 Jan;11(1):66-80. [doi: 10.1123/japa.11.1.66]
11. Thomas JC, Odonkor C, Griffith L, Holt N, Percac-Lima S, Leveille S, et al. Reconceptualizing balance: attributes associated with balance performance. Exp Gerontol 2014 Sep;57:218-223 [FREE Full text] [doi: 10.1016/j.exger.2014.06.012] [Medline: 24952097]
12. Kojima G, Masud T, Kendrick D, Morris R, Gawler S, Treml J, et al. Does the timed up and go test predict future falls among British community-dwelling older people? Prospective cohort study nested within a randomised controlled trial. BMC Geriatr 2015 Apr 03;15(1):38 [FREE Full text] [doi: 10.1186/s12877-015-0039-7] [Medline: 25887660]
13. Gillespie LD, Robertson MC, Gillespie WJ, Sherrington C, Gates S, Clemson LM, et al. Interventions for preventing falls in older people living in the community. Cochrane Database Syst Rev 2012 Sep 12(9):CD007146. [doi: 10.1002/14651858.CD007146.pub3] [Medline: 22972103]
14. Renfro M, Bainbridge DB, Smith ML. Validation of evidence-based fall prevention programs for adults with intellectual and/or developmental disorders: a modified Otago exercise program. Front Public Health 2016;4:261 [FREE Full text] [doi: 10.3389/fpubh.2016.00261] [Medline: 27999771]
15. Campbell AJ, Robertson MC, Gardner MM, Norton RN, Buchner DM. Falls prevention over 2 years: a randomized controlled trial in women 80 years and older. Age Ageing 1999 Oct;28(6):513-518. [doi: 10.1093/ageing/28.6.513] [Medline: 10604501]
16. Campbell AJ, Robertson MC, Gardner MM, Norton RN, Tilyard MW, Buchner DM. Randomised controlled trial of a general practice programme of home-based exercise to prevent falls in elderly women. BMJ 1997 Oct 25;315(7115):1065-1069 [FREE Full text] [doi: 10.1136/bmj.315.7115.1065] [Medline: 9366737]
17. Campbell AJ, Robertson MC, La Grow SJ, Kerse NM, Sanderson GF, Jacobs RJ, et al. Randomised controlled trial of prevention of falls in people aged > or =75 with severe visual impairment: the VIP trial. BMJ 2005 Oct 08;331(7520):817 [FREE Full text] [doi: 10.1136/bmj.38601.447731.55] [Medline: 16183652]
18. Robertson MC, Devlin N, Gardner MM, Campbell AJ. Effectiveness and economic evaluation of a nurse delivered home exercise programme to prevent falls: randomised controlled trial. BMJ 2001 Mar 24;322(7288):697-701 [FREE Full text] [doi: 10.1136/bmj.322.7288.697] [Medline: 11264206]
19. Robertson MC, Gardner MM, Devlin N, McGee R, Campbell AJ. Effectiveness and economic evaluation of a nurse delivered home exercise programme to prevent falls: controlled trial in multiple centres. BMJ 2001 Mar 24;322(7288):701-714 [FREE Full text] [doi: 10.1136/bmj.322.7288.701] [Medline: 11264207]
20. Stevens J, Burns E. A CDC compendium of effective fall interventions: what works for community-dwelling older adults. Atlanta: Centers for Disease Control and Prevention; 2015. URL: https://www.cdc.gov/homeandrecreationalsafety/pdf/falls/cdc_falls_compendium-2015-a.pdf [accessed 2021-07-19]
21. Thomas S, Mackintosh S, Halbert J. Does the 'Otago exercise programme' reduce mortality and falls in older adults?: a systematic review and meta-analysis. Age Ageing 2010 Nov;39(6):681-687. [doi: 10.1093/ageing/afq102] [Medline: 20817938]
22. Jessen JD, Lund HH. Study protocol: effect of playful training on functional abilities of older adults: a randomized controlled trial. BMC Geriatr 2017 Jan 19;17(1):27 [FREE Full text] [doi: 10.1186/s12877-017-0416-5] [Medline: 28103811]
23. Valenzuela T, Okubo Y, Woodbury A, Lord SR, Delbaere K. Adherence to technology-based exercise programs in older adults: a systematic review. J Geriatr Phys Ther 2016 Jun 29;41(1):49-69. [doi: 10.1519/JPT.0000000000000095] [Medline: 27362526]
24. Hilton CE. The importance of pretesting questionnaires: a field research example of cognitive pretesting the Exercise referral Quality of Life Scale (ER-QLS). Int J Soc Res Methodol 2015 Oct 07;20(1):21-34. [doi: 10.1080/13645579.2015.1091640]
25. Survey Research Center. Guidelines for Best Practice in Cross-Cultural Surveys, 4th Edition. Ann Arbor: Institute for Social Research, University of Michigan; 2016.
26. Morgado J, Rocha C, Maruta C, Guerreiro M, Martins I. New normative values of Mini Mental States Examination. Sinapse 2009;9(2):10-16.
27. Bohannon RW, Bubela DJ, Magasi SR, Wang Y, Gershon RC. Sit-to-stand test: performance and determinants across the age-span. Isokinet Exerc Sci 2010;18(4):235-240 [FREE Full text] [doi: 10.3233/IES-2010-0389] [Medline: 25598584]

28. Wretenberg P, Arborelius UP. Power and work produced in different leg muscle groups when rising from a chair. Eur J Appl Physiol Occup Physiol 1994;68(5):413-417. [doi: 10.1007/BF00843738] [Medline: 8076621]
29. Silva J, Oliveira E, Moreira D, Nunes F, Caic M, Madureira J, et al. Design and evaluation of a fall prevention multiplayer game for senior care centres. In: Clua E, Roque L, Lugmayr A, Tuomi P, editors. Entertainment Computing. Cham: Springer; 2018:103-114.
30. Jones CJ, Rikli RE, Beam WC. A 30-s chair-stand test as a measure of lower body strength in community-residing older adults. Res Q Exerc Sport 1999 Jun;70(2):113-119. [doi: 10.1080/02701367.1999.10608028] [Medline: 10380242]
31. Podsiadlo D, Richardson S. J Am Geriatr Soc 1991 Feb;39(2):142-148. [doi: 10.1111/j.1532-5415.1991.tb01616.x] [Medline: 1991946]
32. Branco PS. Determinação dos Pontos de Corte para Elevado Risco de Queda e Mobilidade Normal da Versão Portuguesa da Activities-Specific Balance Confidence (ABC) Scale. Revista da Sociedade Portuguesa de Medicina Física e de Reabilitação 2013;24(2):12-17.
33. Preto L, Mendes E, Novo A, Santos A. Mudanças no desempenho funcional e composição corporal em idosos institucionalizados: um estudo longitudinal. In: Nery de Souza D, Rua M, editors. Cuidadores Informais de Pessoas Idosas: Caminhos de Mudança. Aveiro: Universidade Aveiro Editora; 2013:331-335.
34. Bean JF, Kiely DK, LaRose S, Leveille SG. Which impairments are most associated with high mobility performance in older adults? Implications for a rehabilitation prescription. Arch Phys Med Rehabil 2008 Dec;89(12):2278-2284. [doi: 10.1016/j.apmr.2008.04.029] [Medline: 19061739]
35. Jette AM. Disablement outcomes in geriatric rehabilitation. Med Care 1997 Jun;35(6 Suppl):JS28-JS44. [doi: 10.1097/00005650-199706001-00005] [Medline: 9191712]
36. Winograd CH, Lemsky CM, Nevitt MC, Nordstrom TM, Stewart AL, Miller CJ, et al. Development of a physical performance and mobility examination. J Am Geriatr Soc 1994 Jul;42(7):743-749. [doi: 10.1111/j.1532-5415.1994.tb06535.x] [Medline: 8014350]
37. National Center for Injury Prevention and Control. Tools to Implement the Otago Exercice Program: A Program to Reduce Falls, First edition. Atlanta: Division of Unintentional Injury Prevention, Centers for Disease Control and Prevention; 2017. URL: https://www.med.unc.edu/aging/cgwep/files/2018/09/ImplementationGuideforPT.pdf [accessed 2022-06-01]
38. Graf C. The Lawton Instrumental Activities of Daily Living (IADL) Scale. Best Practices in Nursing Care to Older Adults. 2017. URL: https://dementiasherpa.com/wp-content/uploads/2017/09/Lawton-IADL-scale.pdf [accessed 2021-07-19]
39. Mlinac ME, Feng MC. Assessment of activities of daily living, self-care, and independence. Arch Clin Neuropsychol 2016 Sep;31(6):506-516. [doi: 10.1093/arclin/acw049] [Medline: 27475282]
40. Peng L, Lu W, Liang C, Chou M, Chung C, Tsai S, Taiwan Stroke Postacute Care (PAC) Study Group. Functional outcomes, subsequent healthcare utilization, and mortality of stroke postacute care patients in Taiwan: a nationwide propensity score-matched study. J Am Med Dir Assoc 2017 Nov 01;18(11):990.e7-990.e12. [doi: 10.1016/j.jamda.2017.06.020] [Medline: 28804011]
41. Araújo F, Ribeiro J, Oliveira A, Pinto C, Martins T. 2008 Presented at: Actas do 7th Congresso Nacional de Psicologia da Saúde 2008: Intervenção em Psicologia e Saúde; 2008; Porto. [doi: 10.25248/reas344_2018]
42. Cruz DTD, Duque RO, Leite ICG. Prevalence of fear of falling, in a sample of elderly adults in the community. Rev Bras Geriatr Gerontol 2017 May;20(3):309-318. [doi: 10.1590/1981-22562017020.160176]
43. Delbaere K, Close JCT, Mikolaizak AS, Sachdev PS, Brodaty H, Lord SR. The Falls Efficacy Scale International (FES-I). A comprehensive longitudinal validation study. Age Ageing 2010 Mar;39(2):210-216. [doi: 10.1093/ageing/afp225] [Medline: 20061508]
44. Scheffer AC, Schuurmans MJ, van Dijk N, van der Hooft T, de Rooij SE. Fear of falling: measurement strategy, prevalence, risk factors and consequences among older persons. Age Ageing 2008 Jan;37(1):19-24. [doi: 10.1093/ageing/afm169] [Medline: 18194967]
45. Yardley L, Beyer N, Hauer K, Kempen G, Piot-Ziegler C, Todd C. Development and initial validation of the Falls Efficacy Scale-International (FES-I). Age Ageing 2005 Nov;34(6):614-619 [FREE Full text] [doi: 10.1093/ageing/afi196] [Medline: 16267188]
46. Marques-Vieira CMA, Sousa LMM, Severino S, Sousa L, Caldeira S. Cross-cultural validation of the falls efficacy scale international in elderly: systematic literature review. J Clin Gerontol Geriatr 2016 Sep;7(3):72-76. [doi: 10.1016/j.jcgg.2015.12.002]
47. Figueiredo D, Santos S. Cross-cultural validation of the Falls Efficacy Scale-International (FES-I) in Portuguese community-dwelling older adults. Arch Gerontol Geriatr 2017 Jan;68:168-173. [doi: 10.1016/j.archger.2016.10.010] [Medline: 27810665]
48. Brooke J. SUS: a "quick and dirty" usability scale. In: Usability Evaluation in Industry. London: Taylor & Francis; 1996:189-194.
49. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. Int J Human Comput Interact 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
50. Martins AI, Rosa AF, Queirós A, Silva A, Rocha NP. European Portuguese validation of the System Usability Scale (SUS). Procedia Comput Sci 2015;67:293-300. [doi: 10.1016/j.procs.2015.09.273]

51. Belloni G, Büla C, Santos-Eggimann B, Henchoz Y, Seematter-Bagnoud L. A single question as a screening tool to assess fear of falling in young-old community-dwelling persons. J Am Med Dir Assoc 2020 Sep;21(9):1295-1301. [doi: 10.1016/j.jamda.2020.01.101] [Medline: 32165062]
52. Murphy SL, Dubin JA, Gill TM. The development of fear of falling among community-living older women: predisposing factors and subsequent fall events. J Gerontol A Biol Sci Med Sci 2003 Oct;58(10):M943-M947 [FREE Full text] [doi: 10.1093/gerona/58.10.m943] [Medline: 14570863]
53. Zijlstra GAR, van Haastregt JCM, van Eijk JTM, van Rossum E, Stalenhoef PA, Kempen GIJM. Prevalence and correlates of fear of falling, and associated avoidance of activity in the general population of community-living older people. Age Ageing 2007 May;36(3):304-309. [doi: 10.1093/ageing/afm021] [Medline: 17379605]

Abbreviations

30CST: 30-Second Chair Stand Test
4SBT: 4-Stage Balance Test
AICOS: Center for Assistive Information and Communication Solutions
ESEP: Escola Superior de Enfermagem do Porto
FES-I: Falls Efficacy Scale–International
IADL: Lawton Instrumental Activities of Daily Living Scale
IMU: inertial measurement unit
MMSE: Mini Mental State Examination
OEP: Otago Exercise Program
STEADI: Stopping Elderly Accidents, Deaths, and Injuries
SUS: System Usability Scale
TUG: Timed Up and Go

Edited by A Mavragani; submitted 01.12.20; peer-reviewed by R Boumans, K Liu, Y Cao, M Erdt; comments to author 18.08.21; revised version received 04.11.21; accepted 22.05.22; published 29.06.22.

Please cite as:
Nogueira MN, Silva J, Nogueira I, Pacheco MN, Lopes J, Araújo F
Physical Exercise Program on Fall Prevention Using Technological Interface: Pretest Study
JMIR Form Res 2022;6(6):e26196
URL: https://formative.jmir.org/2022/6/e26196
doi:10.2196/26196
PMID:35767321

Original Paper

Monitoring and Managing Lifestyle Behaviors Using Wearable Activity Trackers: Mixed Methods Study of Views From the Huntington Disease Community

Philippa Morgan-Jones[1], PhD; Annabel Jones[2], MBChB; Monica Busse[3], PhD; Laura Mills[3], BSc; Philip Pallmann[3], PhD; Cheney Drew[3], PhD; Astri Arnesen[4], BSc; Fiona Wood[2], PhD; Domino-HD Consortium[5]

[1]School of Engineering, Cardiff University, Cardiff, United Kingdom
[2]Division of Population Medicine, Cardiff University, Cardiff, United Kingdom
[3]Centre for Trials Research, Cardiff University, Cardiff, United Kingdom
[4]European Huntington Association, Oslo, Norway
[5]see Acknowledgments

Corresponding Author:
Fiona Wood, PhD
Division of Population Medicine
Cardiff University
8th floor, Neuadd Meirionnydd
University Hospital of Wales
Cardiff, CF14 4YS
United Kingdom
Phone: 44 2920687185
Email: wood@cardiff.ac.uk

Abstract

Background: There are early indications that lifestyle behaviors, specifically physical activity and sleep, may be associated with the onset and progression of Huntington disease (HD). Wearable activity trackers offer an exciting opportunity to collect long-term activity data to further investigate the role of lifestyle, physical activity, and sleep in disease modification. Given how wearable devices rely on user acceptance and long-term adoption, it is important to understand users' perspectives on how acceptable any device might be and how users might engage over the longer term.

Objective: This study aimed to explore the perceptions, motivators, and potential barriers relating to the adoption of wearable activity trackers by people with HD for monitoring and managing their lifestyle and sleep. This information intended to guide the selection of wearable activity trackers for use in a longitudinal observational clinical study.

Methods: We conducted a mixed methods study; this allowed us to draw on the potential strengths of both quantitative and qualitative methods. Opportunistic participant recruitment occurred at 4 Huntington's Disease Association meetings, including 1 international meeting and 3 United Kingdom–based regional meetings. Individuals with HD, their family members, and carers were invited to complete a user acceptance questionnaire and participate in a focus group discussion. The questionnaire consisted of 35 items across 8 domains using a 0 to 4 Likert scale, along with some additional demographic questions. Average questionnaire responses were recorded as positive (score>2.5), negative (score<1.5), or neutral (score between 1.5 and 2.5) opinions for each domain. Differences owing to demographics were explored using the Kruskal-Wallis and Wilcoxon rank sum tests. Focus group discussions (conducted in English) were driven by a topic guide, a vignette scenario, and an item ranking exercise. The discussions were audio recorded and then analyzed using thematic analysis.

Results: A total of 105 completed questionnaires were analyzed (47 people with HD and 58 family members or carers). All sections of the questionnaire produced median scores >2.5, indicating a tendency toward positive opinions on wearable activity trackers, such as the devices being advantageous, easy and enjoyable to use, and compatible with lifestyle and users being able to understand the information from trackers and willing to wear them. People with HD reported a more positive attitude toward wearable activity trackers than their family members or caregivers (P=.02). A total of 15 participants participated in 3 focus groups. Device compatibility and accuracy, data security, impact on relationships, and the ability to monitor and self-manage lifestyle behaviors have emerged as important considerations in device use and user preferences.

Conclusions: Although wearable activity trackers were broadly recognized as acceptable for both monitoring and management, various aspects of device design and functionality must be considered to promote acceptance in this clinical cohort.

(JMIR Form Res 2022;6(6):e36870) doi:10.2196/36870

KEYWORDS

Huntington disease; activity tracker; perceptions; digital technologies; physical activity; qualitative research; survey

Introduction

Background

Huntington disease (HD) is a hereditary degenerative neurological disease that affects between 6 and 13 people per 100,000 in the general population [1]. The disease is characterized by the complex presentation of clinical symptoms involving motor, cognitive, and behavioral impairments [2]. People with HD experience a progressive decline in their quality of life and function over 15 to 20 years and premature death [2].

There are indications that lifestyle behaviors such as physical activity and sleep may be linked to the onset and progression of HD [3], with a systematic review indicating preliminary support for the benefits of exercise and physical activity in HD [4] and the European Huntington's Disease Network producing a Physiotherapy Guidance Document for HD [5] supporting physical activity. Sleep disturbance is another feature of HD [6]; however, it is difficult to obtain objective measures of physical activity and sleep that are clinically relevant and valid. Previous studies have relied on subjective reporting from patients, which can be subject to recall bias, high levels of missing data, and can take substantial time to complete, leading to significant measurement bias in the results [7].

Wearable activity trackers may provide a suitable platform to objectively capture physical activity and sleep data. There has been a surge in the availability of wearable digital technology in the consumer market for measuring daily activities and lifestyle habits [8]. Short-term goal setting and instant feedback abilities make them efficacious motivational tools to promote health-related behaviors, and they are increasingly used to facilitate the management of some long-term conditions [9,10].

Consumer engagement with these devices is complex [11]. Although some people may hold a largely positive attitude toward digital wearable devices, others may regard them as invasive or intrusive. Adherence (ie, wearing the device) can be an issue, with some clinical research studies showing that the device was only worn for approximately 50% of the time over 2 days [12,13]. Concerns have also been raised regarding the validity and reliability of the recorded data [14]. Therefore, researchers must consider the balance between clinical validity and consumer compatibility when selecting an appropriate activity tracker device for research and clinical use.

Objective

The Multi-Domain Lifestyle Targets for Improving Huntington Disease (DOMINO-HD) study is a multinational observational study that aims to explore the interplay between lifestyle and genetic factors and HD outcomes. The study hypothesizes that the prognosis of HD can be influenced by the use of digital technology to detect symptoms and allow modification of lifestyle factors. The aim of the DOMINO-HD substudy reported here is to gain an understanding of the views of individuals with HD and their families or caregivers on digital wearable activity tracker devices that will be used in the wider DOMINO-HD research program. It is vital to explore the experiences and preferences of patients with HD and their caregivers regarding wearable activity trackers to maximize downstream engagement with the technology in the DOMINO-HD study.

Methods

Study Design

A mixed methods approach was used by combining self- or proxy-completed questionnaires and focus group discussions. Using mixed methods research allowed us to integrate both types of data to seek a wider range of attitudes [15]. Our questionnaire data allowed us to assess the extent of support for activity trackers in relation to predetermined questions. Our focus group data allowed us to gain insights into participants' explanations of their views, decision-making processes, and previous experiences. In designing our research, we drew on the Technology Acceptance Model, which examines how users come to accept and use a technology [16].

Recruitment

Opportunistic recruitment took place from September to November 2019 at Huntington's Disease Association meetings. These included the patient-centered 2019 European Huntington's Association (EHA) meeting in Bucharest, Romania, and regional Huntington's Disease Association meetings for patients and family members across the United Kingdom (Belfast, Northern Ireland; Cardiff, Wales; and Telford, England). Members of the DOMINO-HD research team attended these meetings, distributing promotional materials and inviting people to participate in both the questionnaire and focus group aspects of this study.

The inclusion criteria to participate in both the questionnaire and focus groups included being aged ≥18 years and having a genetically confirmed diagnosis of HD or being a family member or carer for someone who has HD. This facilitated the collation of opinions across a range of stakeholders within the HD community.

As no personal or sensitive data were requested within the questionnaire, potential participants were made aware that completion of the voluntary questionnaire would be regarded as providing their consent to participate in the study. Those participating in the focus groups were provided with an information sheet and asked to provide written informed consent

before participating. As the focus groups were conducted during the conference, participants may have had only a few hours to consider whether they wished to participate, but the facilitator was mindful of explaining that they could withdraw if they wished. Furthermore, no sensitive topics or clinical information were discussed and shared during the focus groups.

Questionnaire Design

Participants were asked to complete a questionnaire on their thoughts on wearable technology that can monitor lifestyle behaviors (Multimedia Appendix 1). More specifically, an adapted version of the questionnaire developed by Wu et al [17] was used; the questionnaire had originally been designed to explore consumers' intention to accept a smartwatch, with a recognized potential to be extended to other wearable device studies. The modifications made were based on feedback from a Patient and Public Involvement workshop and related to the layout of the questionnaire (visualizing the Likert scale with colored smiley faces of varying happiness and reducing the number of questions per page). Slight alterations to the wording of section 1 of the Wu et al [17] questionnaire were made to reflect the potential benefits of using activity trackers within the HD community.

A total of 39 questions fell within eight subsections: relative advantage, ease of use, compatibility, result demonstrability, enjoyment, social influence, attitude, and behavioral intention (refer to Table 1 for each section descriptor). Each question within these subsections was measured using a 5-point Likert scale based on negative and positive anchors, ranging from 0=*strongly disagree* to 4=*strongly agree*. An additional option, *I don't know*, was also available. Two additional questions were asked to identify the participants' association with HD (ie, whether they had a genetically confirmed diagnosis of HD or were a family member or carer of someone with HD) and to determine their age group (≤24 years, 25-34 years, 35-44 years, 45-54 years, and ≥55 years).

Table 1. Questionnaire domains, descriptors, and example statements.

Questionnaire domain	Section descriptor	Example statement
1. Relative advantage	Whether wearable activity trackers are perceived to be beneficial to those with HD[a]	"The activity tracker would help me to monitor my physical activity and sleep."
2. Ease of use	Whether people with HD would find wearable activity trackers easy to use	"I believe that the activity tracker would be easy to use."
3. Compatibility	Whether activity trackers would be compatible with people who have HD	"An activity tracker is something that I can see fitting my current habits."
4. Result demonstrability	Whether people with HD would be able to understand the information from a wearable activity tracker and be able to explain it to others	"Observing how I do things differently before and after using an activity tracker would be easy for me."
5. Enjoyment	Whether people with HD would enjoy using a wearable activity tracker	"Using an activity tracker would be an ideal recreation"
6. Social influence	Whether people with HD would find wearable activity trackers helpful in raising their social status and whether they value the opinion of others on whether they use such a device	"Anyone who uses a fitness tracker would have higher social status within my social circle."
7. Attitude	Whether people with HD have a positive attitude toward wearing a wearable activity tracker	"Using an activity tracker would be a positive decision."
8. Behavioral intention	Whether people with HD would be willing to use a wearable activity tracker	"I would be willing to use an activity tracker."

[a]HD: Huntington disease.

Participants were able to complete the questionnaire on paper or electronically (via the Bristol Online Surveys platform), which took approximately 10 to 15 minutes to complete. The participants were alerted to this study via a flyer in their delegate packs, which contained a QR code with a link to the questionnaire. A member of the research team was always present at a stand at the conference, allowing delegates the opportunity to complete the questionnaire on an iPad. The paper versions were also handed out and then returned to the researchers at the stand. Given that the European participants attended the EHA Bucharest event, the questionnaire was made available in English, Polish, German, and Spanish, with translations from English being made by DOMINO-HD partners.

Focus Groups

Each focus group was conducted in English, involved 4 to 6 participants, and was facilitated by a researcher trained and experienced in the method and research related to patients with HD. In addition, an additional observer from the research team was present to support the group and make notes. Owing to the relatively narrow aims of the focus groups, we anticipated that we would need to conduct 2 or 3 focus groups to achieve data saturation. They lasted approximately 1 hour (although timings were flexible to the needs and wishes of the participants involved in each group) and were audio recorded using an encrypted audio recorder. They were conducted in a private room during the conference. A semistructured topic guide, developed to explore patients' attitudes toward wearable devices in osteoarthritis [18], was modified for relevance and used to

guide the topics (Multimedia Appendix 2). In addition, short focusing activities were conducted, including a group ranking in which participants were asked to collectively rank in order of importance the features of activity trackers. The main purpose of the ranking exercise was to facilitate discussion without interpreting it quantitatively. Finally, participants were asked to discuss a vignette scenario around wearable technology (Multimedia Appendix 3). The list of items to be ranked and the vignette were developed by the research team with input from the DOMINO-HD Patient and Public Involvement group. The research team was from a range of backgrounds, including physiotherapy (research and clinical), engineering, medicine, and sociology. Our multidisciplinary background encouraged us to explain and keep checking our various assumptions and biases during data collection and analysis.

Data Analysis

All questionnaire responses were entered into the digital version of the questionnaire, where responses were coded and exported using Microsoft Excel. The data were then imported into Spyder (Python 3.8; Python Software Foundation) for further analysis using Pandas [19] and NumPy [20] Python packages. Missing and *I don't know* responses were recoded as *not a number* before the responses to questions within each section of the questionnaire were averaged for each participant. Participants were required to respond to a minimum of 1 question within a given section of the questionnaire to be included in the analysis for that section.

For each questionnaire subsection, responders were divided into those with a positive opinion (ie, producing a mean section score of >2.5), neutral opinion (ie, producing a mean section score of 1.5-2.5), or negative opinion (ie, producing a mean section score of <1.5) and reported as a percentage of the total number of responders for each subsection. The median cohort response for each subsection was also reported alongside the IQR and similarly interpreted in relation to positive, neutral, or negative opinions. A more elaborate quantitative analysis of the association between responses from each questionnaire subsection was not possible as part of this study because of the nature of the responses collated (limited variance across participants and/or questionnaire subsections).

A Wilcoxon rank sum test was performed to determine whether questionnaire responses differed between those with a confirmed diagnosis of HD and those who were family members or carers for someone with HD (*a priori* level of significance set to P=.05, with P value Holm correction performed for multiple comparisons). A Kruskal-Wallis test was performed to determine whether questionnaire responses differed between the 5 age categories (*a priori* level of significance was set to P=.05, with P value Holm correction performed for multiple comparisons).

All focus groups were audio recorded and transcribed verbatim by a professional transcription agency. Transcriptions were read and checked for errors by a researcher who was present in all focus groups, and the primary identifiers were deidentified. An exploratory thematic analysis [21] was performed by a single researcher using the NVivo 12 software (QSR International). After conducting 2 focus groups, the team reflected on the data saturation and determined that a third focus group would be beneficial. Following the additional analysis of these data, the team deemed that we had reached a point of data saturation. The topic themes allocated to nodes were mutually agreed upon across the broader research team and refined during the coding process as topics were unearthed. A third of the data were double coded by another researcher, and then coding decisions were discussed to encourage a more explicit engagement with the data and check coding consistency [22].

The themes identified through the focus group thematic analysis were used to provide context for the questionnaire responses and explore the underpinning opinions of the HD community toward wearable activity trackers.

Ethics Approval

Ethics approval was obtained from the Cardiff University School of Medicine Ethics Committee (ref 19/71, September 19, 2019).

Results

Questionnaire Results

A total of 114 participants completed the study questionnaire, of which 9 (7.8%) were removed from the data set owing to incomplete data (n=1) or failing to fall within the demographics of interest (having a genetically confirmed diagnosis of HD or being a family member or carer of someone with HD [n=8]). This left a questionnaire cohort size of 105 with recorded demographics, as described in Table 2.

The participants were overwhelmingly positive regarding the use of wearable activity trackers (Table 3). All sections of the questionnaire produced a median cohort response >2.5, representing a tendency for positive opinions toward the use of wearable activity trackers (such as devices being advantageous, easy and enjoyable to use, and compatible with one's lifestyle and users being able to understand the information from wearable activity trackers and willing to wear them).

The only differences observed were in the behavioral intention questionnaire response score because of age (Table 4), with the oldest age group (≥55 years) having a lower median response score than the two youngest age categories (≤24 years and 25-34 years). However, pairwise comparisons were not statistically different when P values were adjusted for multiple comparisons (adjusted P>.05). Responders who had a genetic diagnosis of HD were found to have a significantly more positive median response for the ease of use, enjoyment, attitude, and behavioral intention sections when compared with those who cared for or were a family member of someone with HD. However, this only remained statistically significant for the attitude section, following the Holm correction procedure (adjusted P<.05).

Table 2. Questionnaire participant demographics (N=105).

	Participants, n (%)
Age group (years)	
≤24	7 (6.7)
25-34	18 (17.1)
35-44	23 (21.9)
45-54	20 (19)
≥55	35 (33.3)
Not specified	2 (1.9)
Association with HD[a]	
Having a genetically confirmed diagnosis of HD	47 (44.8)
Being a family member or carer for a person with HD	58 (55.2)
Language of questionnaire completion	
English	93 (88.6)
Spanish	3 (2.9)
German	5 (4.8)
Polish	4 (3.8)

[a]HD: Huntington disease.

Table 3. The percentage of median positive, neutral, and negative responses along with the total number of responders for each questionnaire section[a].

Questionnaire section	Positive responses[b], %	Neutral responses[c], %	Negative responses[d], %	Cohort response, median (IQR; range)	Total number of responders (n)
Relative advantage	92.23	4.85	2.91	3.6 (3.0-4.0; 0-4)	103
Ease of use	81.55	14.56	3.88	3.4 (2.8-4.0; 0-4)	103
Compatibility	65.69	23.53	10.78	3.3 (2.3-4.0; 0-4)	102
Result demonstrability	78.43	17.65	3.92	3.0 (2.8-4.0; 0-4)	102
Enjoyment	60.61	27.27	12.12	3.0 (2.0-4.0; 0-4)	99
Social influence	53.92	34.31	11.76	2.6 (2.0-4.0; 0-4)	102
Attitude	74.26	18.81	6.93	3.2 (2.5-4.0; 0-4)	101
Behavioral intention	88.35	7.77	3.88	3.7 (3.0-4.0; 0-4)	103

[a]Participants were asked to rate on a scale of 0 (strongly disagree) to 4 (strongly agree).

[b]Scores>2.5.

[c]Scores between 1.5 and 2.5.

[d]Scores<1.5.

Table 4. Median and IQR for cohort-level questionnaire response scores and results when testing for differences in questionnaire response based on age categories (Kruskal-Wallis test) and whether they were a carer or patient (Wilcoxon rank sum test).

Questionnaire section	Kruskal-Wallis results			Wilcoxon rank sum results		
	Test statistic	*P* value	Adjusted *P* value	Test statistic	*P* value	Adjusted *P* value
Relative advantage	6.195	.19	.79	1.996	.046	.23
Ease of use	6.486	.17	.79	2.343	.02[a]	.12
Compatibility	6.61	.16	.79	1.07	.29	.57
Result demonstrability	7.87	.10	.67	1.558	.12	.36
Enjoyment	5.864	.21	.79	1.986	.047[a]	.23
Social influence	5.234	.26	.79	0.802	.42	.57
Attitude	7.903	.10	.67	3.073	.002[a]	.02[a]
Behavioral intention	9.773	.04[a]	.35	2.602	.009[a]	.07

[a]Statistically significant *P* values (at the 5% level).

Focus Group Results

Overview

A total of three focus groups were conducted: 2 in Bucharest at the EHA patient conference (involving 5 participants in the first focus group and 6 participants in the second), and 1 in Cardiff at the patient HD meeting (involving 4 participants). The demographics of the focus group are presented in Table 5.

The results of the group ranking exercise undertaken during the focus groups are shown in Table 6. Although there were obvious differences in how each focus group ranked what was important to them, the features that appeared to be most important were accuracy, comfort, and ease of use. Appearance and ability to use the watch in other aspects of their lives were the least important.

A total of 4 overarching themes were developed to describe the acceptability of wearable activity trackers to people with HD. These included the accessibility and compatibility of a device, its impact on a person's relationships, whether it can be used effectively for self-management and monitoring of lifestyle behaviors, and the security of the data being collected. We discuss each of these themes in turn with illustrative extracts from the focus groups, although we have withheld participant characteristics from the quote to prevent deductive disclosure because of concerns about the HD community being relatively small.

Table 5. Focus group participant demographics (N=15).

	Cohort sample size, n (%)
Gender	
Male	5 (33)
Female	10 (67)
Experience with wearable devices	
Yes	5 (33)
No	10 (37)
Genetic status	
Having a genetically confirmed diagnosis of HD[a]	5 (33)
Not disclosed	10 (67)
Country of residence[b]	
United Kingdom	7 (46)
Other European country	4 (27)
Other	4 (27)

[a]HD: Huntington disease.

[b]Categories have been collapsed for participants' stated country of residence owing to small numbers.

Table 6. Results of the ranking exercise in which participants were asked to rank the most important factors that would determine their engagement with a wearable device.

Rank	Group 1	Group 2	Group 3
1	Ease of use	Accuracy	Accuracy
2	Getting feedback from the device	Cost	Comfort
3	Comfort	Comfort	Ease of use
4	Keeping data safe and secure	Battery life	Where the tracker is located on the body
5	Accuracy	Keeping data safe and secure	Keeping data safe and secure
6	Battery life	Obtaining feedback from the device	Obtaining feedback from the device
7	Where the tracker is located on the body	Ease of use	Cost
8	Cost	Where the tracker is located on the body	Battery life
9	Being able to use the watch for things not related to the study	Being able to use the watch for things not related to the study	Appearance
10	Appearance	Appearance	Being able to use the watch for things not related to the study

Theme 1: Accessibility and Compatibility

Focus group participants acknowledged that wearable devices must be easy to use for people with HD to engage with them. This was deemed particularly important as this cohort was thought to have a wide-ranging level of digital technology experience. However, the participants were clear that they did not want an easy-to-use device at the expense of limited functionality. Rather, they were clear that they valued the special features of many commercially available activity watches:

> *When you hit 10,000 steps and you've got it on your wrist there's these like fireworks go off...and it vibrates and whatever and you know, you're wonderful and then on the actual app itself, like on the phone, erm, that it's synced to, it's, like I thought it was really clever.*

All participants agreed that battery life was an important factor when considering if a device was easy to use. It was acknowledged that although charging a device was a necessity, a short battery life was not synonymous with ease of use. Those who already owned a wearable device noted how often and for how long charging was required:

> *The charge takes about two hours, this, I, I, I done a lot of research and for the money, erm, and for all round what it does, this ticks all the boxes for me.*

It was also felt that battery life diminished over time, and devices often failed to reach the battery lengths advertised:

> *You know the Fitbits say they're supposed to last five days, they don't last five days, no way they last five days. I charge mine every night.*

Concerns were raised as to whether people with HD, particularly as the disease progresses, would be able to remember to remove a device, charge it, and put it back on. This process needed to be simplified and as infrequent as possible. A device that is waterproof and, therefore, reduces the need to take off and put back on again was similarly deemed beneficial. It was also recognized that a simple watch strap would be required so that those with progressive motor symptoms would be able to wear and remove the device. One participant mentioned using different types of straps with their current device, including a magnetic strap that easily clicks together.

The preferred location for a wearable activity tracker was the wrist; however, interest was also shown in devices that could be worn under clothes so that they could be worn more discreetly. Comfort was believed to be a key factor in device adoption, with participants wanting a device that was as unobtrusive as possible and not bulky or large. However, it was recognized that if there was a need and benefit from using a device, then appearance was not a critical factor in device adoption:

> *If you have the need, it looks like whatever it looks like.*

Several participants displayed a reluctance to wearing a device at night, owing to continuous monitoring feeling too invasive and onerous or simply because they do not wear a watch to bed. Although others did accept wearing a device at night for the purpose of research, the overall consensus suggested that nocturnal monitoring with a wearable activity tracker is not compatible in a real-world context:

> *Ultimately to be a part of somebody's medical care then it has to be as less invasive as possible, okay? Because I don't sleep with watches, I'm not going to enjoy sleeping with watches.*

Cost was believed to be a limiting factor; although some were prepared to pay for a high-end device if it met their requirements, others noted that some HD families have financial constraints that must be considered.

Participants believed that wearable activity trackers are generally considered compatible with the lifestyle and daily routine of HD families and highlighted how the cohort often uses such technologies already. It was recognized that the adoption of a given device within this clinical cohort required iOS and Android operating system compatibility. Wearable activity trackers were deemed particularly suited to younger people, as

they were seen to fit well with the global trend of sharing information via social media. Participants felt that remote health monitoring at home had considerable future potential and that similar technologies were needed for patients with HD. There was a view that, where possible, the person with HD should be encouraged to engage with the wearable device themselves to promote autonomy and engagement.

Theme 2: Impact on Relationships

The participants discussed the difficulties of striking a balance between using the device as a helpful tool and ensuring that it did not become socially or emotionally consuming. Dependence was felt to lead to unnecessary stress when trying to meet health-related targets:

> *It does seem to me that addiction is a big issue. I mean I've certainly read that that is the case, like some people they do get quite obsessed with targets and, and you know, if they haven't done it, they get quite worked up and stressed...*

Feedback, although essential for motivation and engagement, was also thought to contribute to frustration. Tracking and monitoring may prove overwhelming for some patients, particularly those with psychiatric conditions, as one participant illustrated when discussing their experience in attempting to *track* and help their sibling with prompts:

> *He said, oh all of these numbers, you know, they're driving me crazy, you know, and it was really, it was really freaking him out.*

Monitoring sleep difficulties was also a source of frustration for some participants:

> *Now the sleep one, even though it tells me you're not sleeping, it doesn't help me, it doesn't tell me what I can do...*
>
> *It's inventing more stress as a consequence.*

Gaining pleasure from a device was considered possible. For a patient with HD, gaining a sense of autonomy over how they manage their disease could improve their outlook and subsequent relationships with others:

> *If the individual feels that by doing this they're taking more control of their situation that's got to be good for the outlook and the interaction with people around them.*

However, it was acknowledged that if a carer or family member was to be heavily involved in the use of the device, a patient may lose their sense of autonomy and subsequently be less likely to engage. It may become an irritant and potentially strain the carer-patient relationship:

> *I can imagine, like a carer or a family member that's trying to promote it for the person to do it, it might not be as successful if the person wants to do it themselves and take ownership of it I suppose.*

Nearly all participants thought that wearable devices would improve their interactions with medical professionals. Clinicians could analyze longitudinal data produced by the device and modify treatment plans accordingly. However, some have suggested that these devices may make some clinicians feel that their clinical experience and judgment are being challenged:

> *Some doctors would definitely take the view that you're trying to do their job for them and not like that. Others would welcome the increased data that they can make judgements on, I've run into both.*

Theme 3: Self-management

The participants suggested that wearable activity trackers opened up the possibility of new ways for clinicians, researchers, and patients to diagnose, monitor, and manage HD. The need to improve how patients with HD interact with their health professionals was noted, along with the advantages of generating longitudinal data to assist in clinical decision-making rather than relying on how a patient presents on a particular day in the clinic:

> *It gives data that is continuous, so now you have a sense over time, all the time, not just when the person goes to hospital or goes to clinic.*

Those who had experience of using wearable activity trackers described them as being an integral component of a healthy lifestyle, giving them motivation and encouragement to reach their goals and a way of recording success:

> *It just helps me track my health and fitness goals, really. It keeps me on track and focussed. Erm, yeah, keeps me focussed I think is the main advantage for me.*

There was a strong desire to receive feedback from a wearable activity tracker to increase personal motivation, allowing the visualization of goals:

> *I will make that extra effort to hit the target if I'm near it.*

However, the participants recognized that not hitting short-term targets may lead to disappointment and subsequent demotivation. For someone with HD who is seeing a decline in function and potentially not meeting their activity targets, this decline in function may be psychologically damaging:

> *If this is tracking disease progression then if Peter [person in the vignette] was to see a significant change in his disease they may, actually may make him more depressed or more concerned and actually be less helpful to him.*

Although many benefits of using wearable activity trackers were noted, there was a strong feeling that these will only be realized if wearable technologies can record measurements with a reasonable level of accuracy (with mention of an 85% accuracy threshold considered as acceptable):

> *If it's not accurate to the degree that it needs to be then there's no point in doing any of this.*
>
> *Let's face it, I mean if it's not accurate it's a complete waste of time.*

The participants also believed that some medical professionals were not very receptive to patients producing their own data. Participants believed that getting clinicians involved in the process of deciding what to measure and how to feed back to

health professionals could improve their acceptance of the technology.

Theme 4: Data Security

Concerns about data security were also raised. It was felt that the misuse of personal data by third parties could lead to potential discrimination and that people with neuropsychiatric disease features may be obsessed with data security. Despite this, people were willing to still engage with a device and share personal information if there was perceived benefit and little risk:

> *Keeping your data safe and secure though would be very important for some people I'd imagine.*

Discussion

The long-term adoption of wearable activity trackers requires users to accept technology in their daily lives. This study sought to understand the potential views and opinions of the HD community toward wearable activity trackers to guide the selection of a wearable activity tracker for use in a longitudinal observational clinical study.

Principal Findings

The responses to wearable devices in the context of HD were largely positive in both the questionnaires and focus groups. Activity trackers need to be accurate, have a useful purpose, be easy and enjoyable to use, and be compatible with the wearer's lifestyle. It was also considered important that this patient group would be able to understand the information received from the tracker. Older participants were less likely to indicate that they would be willing to use a wearable device than younger participants (behavioral intentions). Our focus group data indicated that engagement with the device was facilitated by accuracy, ease of use, comfort, usable feedback, and reliable battery life. The questionnaire data showed that the domain of social influence, or how the activity tracker might affect how other people view the user, was less important. Again, this was paralleled in the focus group data, where the appearance of the device was deemed to be less important in the ranking exercise. Concerns were also raised regarding the suitability of using a wearable device for some patients with HD, particularly those with significant neuropsychiatric symptoms. Reassurances regarding data security also need to be considered.

Findings in the Context of Other Literature

The Technology Acceptancy Model suggests that when users are presented with a new technology, the main factors that influence their intention to use it are perceived usefulness and perceived ease of use [16]. Our data indicate that activity trackers need to have a useful purpose. Usefulness was associated with accuracy in attaining benefits. However, high thresholds of accuracy, suggested by one of our focus group participants, are unrealistic, particularly because of the uncertainty of the accuracy and reliability of data from commercial devices [23]. The expectation of having an accurate wearable activity tracker is consistent with several other studies [24]. Although no participants in this study directly questioned the accuracy of wearable activity trackers, the algorithms embedded within devices are typically developed using the movement patterns of healthy people, which may lead to higher data inaccuracies when applied to people with HD, who often present with motor symptoms involving their upper limbs and pathological gait characteristics. A recent review of wearable activity tracker research [25] found a substantial increase in the literature focusing on wearable activity trackers in recent years, with the most common research theme focusing on concerns about the reliability, accuracy, and validity of the technology itself.

Research has also suggested that wearable devices worn on the wrist have a greater error rate than those worn on other parts of the body [26]. Despite this, the immediate feedback provided by a digital screen on a wrist-watch device, which is essential for sustained user engagement, indicates that it is the preferred location for many users [26]. Accuracy may not need to be perfect, but it is important and must be sufficient to produce reliable clinical data. Researchers will also need to know more about the inaccuracies of measurement to appreciate the minimal clinically relevant differences between populations.

Acceptability is also linked to the ease of use. A device that was small, discrete, and comfortable was preferred. Most participants were not supportive of a cumbersome device that could lead to patient identification and consequent discrimination, which were concerns shared by other neurodegenerative populations [27,28].

Despite numerous potential benefits, concerns have been expressed regarding the relationship with a device that measures several behaviors in this specific patient population. Most research on wearable devices is primarily focused on healthy populations, and research on clinically specific populations is relatively scarce [18]. Therefore, there is a need to consider the risk of harm in the HD population. Disease features such as depression, anxiety, and cognitive impairment can influence social relationships, including engagement with a device [29,30]. Some users may become dependent on wearable technology and place a disproportionate value on the data produced, with consequences on mood and further engagement [31]. Lack of confidence in technology has also been cited as a barrier to engagement in other neurodegenerative diseases, with some patients expressing stress at the thought of wearing such a device [32]. Others may find it intrusive and uncomfortable [14]. Although compliance may not be an issue, welfare concerns can arise. Therefore, researchers and clinicians must assess the suitability of participants using a device on an individual basis to prevent negative psychological consequences. Devices that do not offer user interaction or feedback may prevent this. However, this study suggests that devices with added functionality can increase patient adherence and engagement. Therefore, it is difficult to balance protecting patient vulnerabilities and maximizing user engagement and adherence by providing feedback.

The reluctance to wear a device at night is also shared by other populations, although the importance of continuous monitoring to obtain high-quality sleep data has been recognized [18,33]. Concerns raised about having to remember to physically charge a wearable device were also shared by patients with Parkinson

disease, who cited difficulties in charging a device as a barrier to user engagement [27], partly owing to limited motor control and the dexterity needed to remove the device. Therefore, the design should aim to preserve the battery life and minimize charging requirements.

Concerns regarding data privacy were also raised in this study. These concerns are echoed in the literature [24,25,33]. Although patients may be open to sharing their personal information for the purpose of clinical benefit, there is a risk of data misuse. Therefore, it is vital to ensure that patient information is protected and that both researchers and industries demonstrate transparency regarding data use [33]. However, the literature suggests an overall risk-benefit consideration that favors engagement with wearable devices [34,35].

Participants' perceived motivation and engagement in health-related behaviors because of wearable device use are shared by other research focusing on exercise and rehabilitation in chronic diseases [36,37]. Behavior change techniques instilled by wearable technology and driven by feedback, such as short-term goal setting, prompts, and reward systems, are key features that drive health surveillance [38]. In the context of long-term conditions, wearable technology may overcome limitations in monitoring patient-reported outcomes and encourage greater patient-clinician cooperation [37]. Consistent and reliable measurements provide an opportunity to identify and treat those with active disease and those who are at risk [38]. This has already been recognized by other neurodegenerative populations, and the use of wearable technology has already been introduced in the management of Parkinson disease [28,39]. The clinical application of wearable devices in a real-world context in HD remains limited, although this study suggests that they would be well received.

Limitations

Although these findings were elicited in the context of HD, we believe that they are applicable to other long-term conditions, particularly those that can be influenced by health-related behaviors. We note that the questionnaire for the survey was not validated for the HD population or any other neurodegenerative population. However, we received feedback from our Patient and Public Involvement group about its relevance and acceptability. Our questionnaires were available only in 4 languages and were not back translated. However, the vast majority of the participants completed the questionnaire in English. Opportunistic sampling from HD meetings was deemed the most convenient recruitment method for this study. A purposeful sample strategy, which sought a balance of gender, age, and nationality of our participants, might have helped the representativeness of the sample but was considered impractical when recruiting at site. Therefore, our findings may not be generalizable to a broader cohort of patients with HD and their caregivers. We opted to collect only brief sociodemographic data on age group and whether the participant was a patient with HD or a caregiver, as we thought these might be the most important differences. However, we acknowledge that we were not able to report participants' views in relation to gender, socioeconomic status, or educational background. In addition, focus group research can result in social desirability bias, leading to overestimation of a positive response [40], but this may have been rebalanced by our mixed methods design, where participants were self- or proxy-completing the questionnaire and thus had limited and anonymous contact with the research team. Functionality issues, fatigue, and cognitive impairment may also have prevented members of the HD population from participating in our research [30,41,42]. Nonetheless, there was flexibility in our design to allow people to choose to participate in either our focus groups or to complete our questionnaire.

Conclusions

Our participants demonstrated a positive attitude toward the application of wearable activity tracker technology. The results of this study guided the design features of the chosen wearable device to be included in the DOMINO-HD study. The selection of and appropriate modifications to a wearable device should maximize user engagement and adherence for home monitoring throughout the observational study period.

As our research population was relatively specific, we recommend that other studies test the acceptability of wearable activity trackers in populations with other chronic conditions. On the basis of the findings of this study, further investigation into an acceptable device measurement schedule and a more detailed investigation into the acceptability of sleep monitoring are also recommended. This study also informed our choice of device that matches patient preferences for the DOMINO-HD longitudinal study, but there is a need to validate that device and measure any inaccuracies to enable a confident estimation of what the minimal detectable change in lifestyle behavior might be.

Acknowledgments

The Multi-Domain Lifestyle Targets for Improving Huntington Disease (DOMINO-HD) study is funded though the European Union Joint Program for Neurodegenerative Disease Research as part of the Joint Program for Neurodegenerative Disease funding call in Health and Social Care (2019), with funding from the Alzheimer's Society, Secretary of State for Health and Social Care, Health and Care Research Wales, Public Health Agency (Northern Ireland), Jacques and Gloria Gossweiler Foundation, Bundesministerium für Bildung und Forschung, Narodowe Centrum Badań i Rozwoju, Swiss National Science Foundation (32ND30_185548), and Health Research Board (JPND-HSC-2018-003). The Centre for Trials Research, Cardiff University, receives infrastructure funding from the Health and Care Research Wales. The authors would like to thank the DOMINO-HD Public and Patient Involvement Group for their important contribution to the development of the questionnaire and the content used within the focus groups.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire.
[PDF File (Adobe PDF File), 708 KB - formative_v6i6e36870_app1.pdf]

Multimedia Appendix 2
Focus group topic guide.
[DOCX File , 15 KB - formative_v6i6e36870_app2.docx]

Multimedia Appendix 3
Vignette for focus group.
[DOCX File , 53 KB - formative_v6i6e36870_app3.docx]

References

1. Rawlins MD, Wexler NS, Wexler AR, Tabrizi SJ, Douglas I, Evans SJ, et al. The prevalence of Huntington's disease. Neuroepidemiology 2016 Jan 30;46(2):144-153 [FREE Full text] [doi: 10.1159/000443738] [Medline: 26824438]
2. Bates G, Tabrizi S, Jones L. Huntington's Disease. Oxford: Oxford University Press; 2014.
3. Mo C, Hannan AJ, Renoir T. Environmental factors as modulators of neurodegeneration: insights from gene-environment interactions in Huntington's disease. Neurosci Biobehav Rev 2015 May;52:178-192. [doi: 10.1016/j.neubiorev.2015.03.003] [Medline: 25770041]
4. Fritz NE, Rao AK, Kegelmeyer D, Kloos A, Busse M, Hartel L, et al. Physical therapy and exercise interventions in Huntington's disease: a mixed methods systematic review. J Huntington Disease 2017 Sep 29;6(3):217-235. [doi: 10.3233/jhd-170260]
5. Quinn L, Busse M. Development of physiotherapy guidance and treatment-based classifications for people with Huntington's disease. Neurodegenerative Disease Manag 2012 Feb;2(1):11-19. [doi: 10.2217/nmt.11.67]
6. Lazar AS, Panin F, Goodman AO, Lazic SE, Lazar ZI, Mason SL, et al. Sleep deficits but no metabolic deficits in premanifest Huntington's disease. Ann Neurol 2015 Oct 21;78(4):630-648 [FREE Full text] [doi: 10.1002/ana.24495] [Medline: 26224419]
7. Althubaiti A. Information bias in health research: definition, pitfalls, and adjustment methods. J Multidiscip Healthc 2016 May;9:211-217 [FREE Full text] [doi: 10.2147/JMDH.S104807] [Medline: 27217764]
8. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 2015 Dec 18;12:159 [FREE Full text] [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]
9. Lupton D. The diverse domains of quantified selves: self-tracking modes and dataveillance. Econ Soc 2016 Apr 15;45(1):101-122. [doi: 10.1080/03085147.2016.1143726]
10. Patel MS, Asch DA, Volpp KG. Wearable devices as facilitators, not drivers, of health behavior change. JAMA 2015 Feb 03;313(5):459-460. [doi: 10.1001/jama.2014.14781] [Medline: 25569175]
11. Lupton D. 'It's made me a lot more aware': a new materialist analysis of health self-tracking. Media Int Aus 2019 Apr 26;171(1):66-79. [doi: 10.1177/1329878x19844042]
12. Matthews CE, Hagströmer M, Pober DM, Bowles HR. Best practices for using physical activity monitors in population-based research. Med Sci Sports Exerc 2012 Jan;44(1 Suppl 1):S68-S76 [FREE Full text] [doi: 10.1249/MSS.0b013e3182399e5b] [Medline: 22157777]
13. Barak S, Wu SS, Dai Y, Duncan PW, Behrman AL, Locomotor Experience Applied Post-Stroke (LEAPS) Investigative Team. Adherence to accelerometry measurement of community ambulation poststroke. Phys Ther 2014 Jan;94(1):101-110 [FREE Full text] [doi: 10.2522/ptj.20120473] [Medline: 24029299]
14. Piwek L, Ellis DA, Andrews S, Joinson A. The rise of consumer health wearables: promises and barriers. PLoS Med 2016 Feb;13(2):e1001953 [FREE Full text] [doi: 10.1371/journal.pmed.1001953] [Medline: 26836780]
15. Shorten A, Smith J. Mixed methods research: expanding the evidence base. Evid Based Nurs 2017 Jul 14;20(3):74-75. [doi: 10.1136/eb-2017-102699] [Medline: 28615184]
16. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319. [doi: 10.2307/249008]
17. Wu L, Wu L, Chang S. Exploring consumers' intention to accept smartwatch. Comput Human Behav 2016 Nov;64:383-392. [doi: 10.1016/j.chb.2016.07.005]
18. Papi E, Belsi A, McGregor AH. A knee monitoring device and the preferences of patients living with osteoarthritis: a qualitative study. BMJ Open 2015 Sep 07;5(9):e007980 [FREE Full text] [doi: 10.1136/bmjopen-2015-007980] [Medline: 26346873]

19. McKinney W. Data Structures for Statistical Computing in Python. In: Proceedings of the Python in Science Conference. 2010 Presented at: Proceedings of the 9th Python in Science Conference; Jun 28 - Jul 3, 2010; Austin, Texas. [doi: 10.25080/majora-92bf1922-00a]
20. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. Nature 2020 Sep 16;585(7825):357-362 [FREE Full text] [doi: 10.1038/s41586-020-2649-2] [Medline: 32939066]
21. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
22. Elliott V. Thinking about the coding process in qualitative data analysis. Qual Report 2018 Nov 24. [doi: 10.46743/2160-3715/2018.3560]
23. Kroll RR, Boyd JG, Maslove DM. Accuracy of a wrist-worn wearable device for monitoring heart rates in hospital inpatients: a prospective observational study. J Med Internet Res 2016 Sep 20;18(9):e253 [FREE Full text] [doi: 10.2196/jmir.6025] [Medline: 27651304]
24. Attig C, Franke T. Abandonment of personal quantification: a review and empirical study investigating reasons for wearable activity tracking attrition. Comput Human Behav 2020 Jan;102:223-237. [doi: 10.1016/j.chb.2019.08.025]
25. Shin G, Jarrahi MH, Fei Y, Karami A, Gafinowitz N, Byun A, et al. Wearable activity trackers, accuracy, adoption, acceptance and health impact: a systematic literature review. J Biomed Inform 2019 May;93:103153 [FREE Full text] [doi: 10.1016/j.jbi.2019.103153] [Medline: 30910623]
26. Fulk GD, Combs SA, Danks KA, Nirider CD, Raja B, Reisman DS. Accuracy of 2 activity monitors in detecting steps in people with stroke and traumatic brain injury. Phys Ther 2014 Feb;94(2):222-229. [doi: 10.2522/ptj.20120525] [Medline: 24052577]
27. Johansson D, Malmgren K, Alt Murphy M. Wearable sensors for clinical applications in epilepsy, Parkinson's disease, and stroke: a mixed-methods systematic review. J Neurol 2018 Aug 9;265(8):1740-1752 [FREE Full text] [doi: 10.1007/s00415-018-8786-y] [Medline: 29427026]
28. Kenny L, O'Riordan C, Wilkinson T, O'Sullivan M, Agurjanovs J, Tedesco S, et al. An exploration of the feasibility and acceptability of wearable technology in Parkinson's disease. Age Ageing 2019;48(3):iii1-ii16. [doi: 10.1093/ageing/afz102.59]
29. Dale M, Cramp R, Shimozaki S. H10 an investigation into gender influence on anxiety and depression in Huntington's Disease (HD) across disease-stage. J Neurol Neurosurgery Psychiatry 2014 Sep 17;85(Suppl 1):A55. [doi: 10.1136/jnnp-2014-309032.155]
30. Unschuld PG, Edden RA, Carass A, Liu X, Shanahan M, Wang X, et al. Brain metabolite alterations and cognitive dysfunction in early Huntington's disease. Mov Disord 2012 Jun;27(7):895-902 [FREE Full text] [doi: 10.1002/mds.25010] [Medline: 22649062]
31. Krantz DS, Baum A, Wideman MV. Assessment of preferences for self-treatment and information in health care. J Pers Soc Psychol 1980 Nov;39(5):977-990. [doi: 10.1037//0022-3514.39.5.977] [Medline: 7441487]
32. Cancela J, Pastorino M, Tzallas AT, Tsipouras MG, Rigas G, Arredondo MT, et al. Wearability assessment of a wearable system for Parkinson's disease remote monitoring based on a body area network of sensors. Sensors (Basel) 2014 Sep 16;14(9):17235-17255 [FREE Full text] [doi: 10.3390/s140917235] [Medline: 25230307]
33. Chiauzzi E, Rodarte C, DasMahapatra P. Patient-centered activity monitoring in the self-management of chronic health conditions. BMC Med 2015 Apr 09;13:77 [FREE Full text] [doi: 10.1186/s12916-015-0319-2] [Medline: 25889598]
34. Li H, Wu J, Gao Y, Shi Y. Examining individuals' adoption of healthcare wearable devices: an empirical study from privacy calculus perspective. Int J Med Inform 2016 Apr;88:8-17. [doi: 10.1016/j.ijmedinf.2015.12.010] [Medline: 26878757]
35. Lancelot Miltgen C, Popovič A, Oliveira T. Determinants of end-user acceptance of biometrics: integrating the “Big 3” of technology acceptance with privacy context. Decision Support Syst 2013 Dec;56:103-114. [doi: 10.1016/j.dss.2013.05.010]
36. Phillips SM, Cadmus-Bertram L, Rosenberg D, Buman MP, Lynch BM. Wearable technology and physical activity in chronic disease: opportunities and challenges. Am J Prev Med 2018 Jan;54(1):144-150 [FREE Full text] [doi: 10.1016/j.amepre.2017.08.015] [Medline: 29122356]
37. Ferriolli E, Skipworth RJ, Hendry P, Scott A, Stensteth J, Dahele M, et al. Physical activity monitoring: a responsive and meaningful patient-centered outcome for surgery, chemotherapy, or radiotherapy? J Pain Symptom Manage 2012 Jun;43(6):1025-1035 [FREE Full text] [doi: 10.1016/j.jpainsymman.2011.06.013] [Medline: 22269181]
38. Lyons EJ, Lewis ZH, Mayrsohn BG, Rowland JL. Behavior change techniques implemented in electronic lifestyle activity monitors: a systematic content analysis. J Med Internet Res 2014 Aug 15;16(8):e192 [FREE Full text] [doi: 10.2196/jmir.3469] [Medline: 25131661]
39. Arora S, Little M, Venkataraman V, Donohue S, Biglan K, Dorsey E, et al. High-accuracy discrimination of Parkinson's disease participants from healthy controls using smartphones. In: Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2014 Presented at: 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); May 4-9, 2014; Florence, Italy. [doi: 10.1109/icassp.2014.6854280]
40. Bergen N, Labonté R. "Everything is perfect, and we have no problems": detecting and limiting social desirability bias in qualitative research. Qual Health Res 2020 Apr;30(5):783-792. [doi: 10.1177/1049732319889354] [Medline: 31830860]
41. Busse ME, Wiles CM, Rosser AE. Mobility and falls in people with Huntington's disease. J Neurol Neurosurg Psychiatry 2009 Jan 16;80(1):88-90 [FREE Full text] [doi: 10.1136/jnnp.2008.147793] [Medline: 19091714]

42. Baker CR, Domínguez DJ, Stout JC, Gabery S, Churchyard A, Chua P, et al. Subjective sleep problems in Huntington's disease: a pilot investigation of the relationship to brain structure, neurocognitive, and neuropsychiatric function. J Neurol Sci 2016 May 15;364:148-153. [doi: 10.1016/j.jns.2016.03.021] [Medline: 27084236]

Abbreviations

DOMINO-HD: Multi-Domain Lifestyle Targets for Improving Huntington Disease
EHA: European Huntington's Association
HD: Huntington disease

Edited by A Mavragani; submitted 28.01.22; peer-reviewed by G Meloghisi, T Holmlund, I Wilson; comments to author 05.04.22; revised version received 18.05.22; accepted 18.05.22; published 29.06.22.

Please cite as:
Morgan-Jones P, Jones A, Busse M, Mills L, Pallmann P, Drew C, Arnesen A, Wood F, Domino-HD Consortium
Monitoring and Managing Lifestyle Behaviors Using Wearable Activity Trackers: Mixed Methods Study of Views From the Huntington Disease Community
JMIR Form Res 2022;6(6):e36870
URL: https://formative.jmir.org/2022/6/e36870
doi:10.2196/36870
PMID:35767346

Original Paper

Experiences of Community Members Engaged in eCPR (Emotional Connecting, Empowering, Revitalizing) Training: Qualitative Focus Group Study

Amanda L Myers[1*], MPH; Mbita Mbao[2*], LICSW; Arya Kadakia[3], BSc; Shira Collings[4*], MS; Karen L Fortuna[5*], PhD, LICSW

[1]Heller School, Brandeis University, Waltham, MA, United States

[2]School of Social Work, Simmons University, Boston, MA, United States

[3]Dartmouth College, Hanover, NH, United States

[4]National Empowerment Center, Lawrence, MA, United States

[5]Department of Psychiatry, Geisel School of Medicine, Dartmouth College, Hanover, NH, United States

[*]these authors contributed equally

Corresponding Author:
Amanda L Myers, MPH
Heller School
Brandeis University
415 South Street
Waltham, MA, 02453
United States
Phone: 1 6038090845
Email: amanda@digitalpeersupport.org

Abstract

Background: The United Nations has called for wide-scale community mental health psychoeducation; however, few programs currently exist. Emotional Connecting, Empowering, Revitalizing (eCPR) is a community education and training program developed by individuals with a lived experience of mental health challenges or trauma. It is designed to provide community members with skills and confidence to support someone experiencing mental health challenges.

Objective: This qualitative study aimed to examine the user experiences of diverse community members engaged in eCPR training. This study reviewed their attitudes toward training and opportunities for improvement in future implementations of training.

Methods: eCPR training participants (N=31) were invited to participate in virtual focus groups between June 2020 and July 2020. Data were analyzed using the rigorous and accelerated data reduction method, which converts raw textual data into concise data tables to develop a codebook, and thematic analysis was performed to identify common themes.

Results: The themes identified when analyzing the data included emotional holding and containment, training feedback, principles and practices of eCPR, implementation, connection in a digital environment, skills practice, and shared experiences.

Conclusions: eCPR may benefit individuals from multiple, diverse demographics. It can enhance their ability to connect with others to understand what it means to be with someone who is experiencing a mental health challenge or crisis, to accept their own emotions, and to be confident in being their most authentic self in both their work and personal lives. eCPR may answer the call of the United Nations by bringing opportunities for authenticity and healing to community settings. Exploring the effects of delivering eCPR in communities on individuals experiencing distress is an important next step. This study found that eCPR may be beneficial to many groups of trainees with varying backgrounds and experiences. These findings are important, as they speak to the potential for eCPR to be implemented in a variety of community settings with the intention of working to improve mental health in everyday settings.

(JMIR Form Res 2022;6(6):e32219) doi:10.2196/32219

KEYWORDS

mental health; trauma; peer support; community mental health education

Introduction

Background

According to the US Department of Health and Human Services' Substance Abuse and Mental Health Services Administration, mental health disorders are "any mental, emotional, or behavioral disorder that can vary in both severity of impact and impairment" [1]. On the basis of their 2019 National Survey on Drug Use and Health, 1 in 5 (20.6%) adults in the United States had a diagnosed mental health disorder [1]. In 2018, nearly 25% of adults aged ≥18 years with mental health challenges reported an unmet need in receiving mental health treatment [2]. More recently, a nationwide survey in the United States estimated that the rates of substance use have risen 13.3% since the onset of the COVID-19 pandemic, anxiety and depressive disorders have increased by 30.9%, and individuals reported increased trauma and stressor-related disorders related to the pandemic by 26.3% [3]. These challenges faced by individuals during a time of social distancing and lockdown measures highlight the need to increase access to and the delivery of mental health support in the community. Delivering services outside of a clinical setting is essential for task shifting purposes that promote individual recovery without the reliance on an understaffed professional clinical task force.

In light of the increase in mental health challenges and trauma owing to the COVID-19 pandemic, coupled with a shortage of practicing mental health professionals, the United Nations released a report "COVID-19 and the Need for Action on Mental Health" in May 2020 [4]. This report called for wide-scale community mental health psychoeducation in which mailpersons, neighbors, Meals on Wheels workers, and other community members can support one another through a pandemic [4]. This global call to action calls for support outside the clinical environment. However, only a few programs of this type exist. The National Empowerment Center, a nonprofit organization led by Oryx Cohen, MPA, and Daniel Fisher, MD, PhD, has developed a web-based training program in community-based mental health interventions that can be practiced among community members in all settings: Emotional Connecting, Empowering, Revitalizing (eCPR).

Promising evidence indicates that eCPR, a community mental health education and training program, provides community members with the skills and confidence to be with someone who is experiencing mental health challenges and/or a mental health crisis [5]. eCPR is a training program developed by individuals with a lived experience of mental health challenges and/or trauma through the National Empowerment Center and has been described in detail in recent literature [5]. eCPR is based on the recovery model of mental health that uses principles of recovery in its approach toward mental health challenges and trauma [6]. Of note, principles suggest that recovery is person driven, is supported through relationships and social networks (such as peers and allies), and emerges from respect and hope. Moreover, the World Health Organization defines recovery from mental health challenges as "gaining and retaining hope, understanding of one's abilities and disabilities, engagement in an active life, personal autonomy, social identity, meaning and purpose in life, and a positive sense of self" [7]. eCPR training uses all these principles outside of clinical environments, specifically, to help individuals in community settings gain insight into mental health challenges and how to provide support to others in their community. A recent study has examined the feasibility and effectiveness of virtually delivered eCPR training. Pre- and posttraining surveys were administered to 560 training participants. A total of 151 participants completed both surveys. Participants demonstrated statistically significant improvements in the ability to identify emotions, communicate nonverbally, share emotions, take care of oneself, feelings of belongingness and connection with others, perceived capacity to support individuals, and symptoms and emotions [5].

Objectives

The purpose of this study is to examine the impact of eCPR on training participants, learn more about their experiences with the training, and evaluate feedback to tailor the implementation of future training. It aims to gain a deeper perspective on the user experience to investigate the underlying mechanisms of the program as well as the priorities and values of various trainees.

Methods

Overview

This study used a peer-academic partnership [8]. Peer-academic partnership is a community-engaged approach that aligns with the 11 principles of community engagement that holds both peers and nonpeer academic researchers as experts, which can contribute to the research process [9]. In this instance, peers who are persons who identify as having a lived experience of a mental health challenge and/or trauma contribute to the research process. This partnership has led to a series of studies [5,10-12]. The peer-academic partnership was used in both the development and implementation of this study by developing the research questions, recruitment, retention, and development of the interview guide (Multimedia Appendix 1); conducting focus groups; and interpreting study findings. Preliminary evidence indicates that using such community engagement methods produces more relevant research, increases engagement, expands the uptake of technologies, and improves clinical outcomes compared with traditional research [13].

Description of eCPR Training

eCPR is designed to be delivered by individuals with lived experiences of mental health challenges including, but not limited to, anxiety, depression, bipolar disorder, schizophrenia, and trauma (ie, peer support specialists and community members) [5] as well as by other community members, such as educators, administrators, and first responders. The 12-hour, 3-day web-based eCPR training included the following modules: (1) connecting with others, (2) using nonverbal communication, (3) cultural empathy across worldviews, (4) learning a trauma-informed approach, (5) addressing feelings of mental distress and thoughts of suicide, (6) empowerment, and (7) revitalization [5]. The aforementioned modules were delivered to the training groups through a variety of teaching methods

including role-play, didactic and conversational instruction, and group experiential learning. eCPR trainers participated in 60 hours of education and training with a postexamination to determine eligibility in delivering the eCPR training [5].

Procedures

Recruitment

Trainers emailed individuals who had participated in the eCPR training to inform them about this opportunity. Participants were recruited using a convenience sample of individuals who expressed interest in participating until the desired sample size was reached. The desired sample size was reached at the point where saturation was achieved, and no additional information was presented [14].

Informed Consent

Before engaging in the focus groups, the participants were provided with an informed consent form via email to sign electronically. The principal investigator (PI) also read the consent statement aloud at the beginning of each focus group. Individuals were given the opportunity to meet with the PI and ask questions pertaining to the study and the informed consent form. Participants were encouraged to contact the PI at any time with further questions and were notified that they could withdraw their participation at any moment.

Demographics

A web-based survey was administered to all participants before eCPR training. Demographic data collected through these surveys were extracted and analyzed for the participants of the focus groups.

Interview Guide

A total of 2 eCPR focus groups occurred virtually via a Health Insurance Portability and Accountability Act–compliant videoconferencing platform between June 2020 and July 2020 and included 31 participants. The authors conducted semistructured focus groups using an interview guide (Multimedia Appendix 1) developed through an iterative design process. Sample questions included, “What are some of the most valuable things you gained from eCPR training?” “How would you compare your eCPR training to other similar trainings you have had in the past?” and “Do you feel eCPR has changed the way you view someone in emotional distress?” Individuals took part in 1-hour focus group sessions to examine the impact of eCPR training on themselves. No incentives were provided for participation in the study.

Ethics Approval

The Institutional Review Board of Dartmouth College approved this study (20201271) and was aligned with the 1968 Declaration of Helsinki rules on research ethics [15].

Statistical Analyses

Descriptive statistics were conducted using SPSS software (version 28.0.0.0; IBM Corp) to describe the demographic characteristics of focus group participants [16]. Focus group data were transcribed and analyzed in Microsoft Excel using the rigorous and accelerated data reduction method for qualitative data coding [17]. This method was chosen for its rigorous team-based approach to organizing qualitative data and streamline analyses. The method converts raw textual data from a general word processing software (Excel) into a more manageable and user-friendly format by producing concise data tables that can be easily reviewed by the team [17]. It was used to generate underlying themes using codes extracted from the qualitative text. The first and second authors performed thematic analysis to identify commonalities in the data-derived codes, which were then translated into the main themes [18]. The authors did not develop a priori codes or hypotheses before coding; instead, codes and themes were developed naturally as they emerged in the review of the data. Codes were assigned to main themes using the *best fit* approach when one or more themes may have been applicable. Furthermore, qualitative data were reviewed by the authors (ALM and MM) to ensure that the results were interpreted for their intended meanings. All results were member checked to maintain the participants’ voices and to ensure that the authors held true to the participants’ viewpoints. Member checking is the process of bringing the authors’ interpretations of the qualitative data back to stakeholders in the population to validate the proper interpretation of the data and resolve any discrepancies that may have emerged [19]. In this case, a qualitative codebook and derived themes were provided to key stakeholders via email. The topics were refined through an iterative process to ensure coherence and adequate data to support each theme.

Results

Participants

The 2 focus groups comprised 31 participants each. Participants were aged ≥25 years and identified themselves as peer support specialists and service users with a lived experience of any mental health condition, recovery coaches, hospital staff leaders, family members, mental health clinicians, nonprofit leaders, and nonprofit workers (Table 1). The inclusion criteria for the focus group participants were individuals (1) who participated in and completed the eCPR training between April 2020 and July 2020; (2) who are aged >18 years; (3) who self-identify as experiencing any mental health condition and/or trauma, family members of individuals with mental health conditions, trauma, or physical health conditions, clinicians, nonprofit leaders and workers, and members of the general community; and (4) who were able to provide consent to participate. Individuals who (1) were aged <18 years, (2) deemed cognitively impaired (defined by inability to complete informed consent independently), or (3) had a designated legal guardian were not eligible to participate in the focus groups.

Table 1. Demographic characteristics of focus group participants (N=31).

Characteristics	Participants, n (%)
Gender	
Male	7 (23)
Female	24 (77)
Age (years)	
18-24	0 (0)
25-34	3 (10)
35-44	3 (10)
45-54	6 (19)
55-64	14 (45)
≥65	5 (16)
Race and ethnicity	
White	23 (74)
Black or African American	4 (13)
American Indian or Alaska Native	1 (3)
More than one race	1 (3)
Hispanic or Latino	2 (7)
Role[a]	
Service user, consumer, or survivor	9 (29)
Peer support specialist	8 (26)
Recovery coach	3 (10)
Family of a person with mental health or substance use issues	8 (26)
Clinician	4 (13)
Administrator	4 (13)
Community member	9 (29)
Work for nonprofit	8 (26)
Other health service provider	4 (13)
Other	8 (26)

[a]Participants were able to select one or more roles.

Demographic data were obtained for all participants in the focus group (N=31). Most participants were female (24/31, 77%) and within the age range of 55-64 years (14/31, 45%), followed by 45-54 years (6/31, 19%). Of the 31 participants, 23 (74%) identified as White, 4 (13%) identified as Black, 2 (7%) identified as Hispanic or Latino, 1 (3%) identified as American Indian or Alaska Native, and 1 (3.2%) identified as belonging to more than one race. When asked about their roles, 29% (9/31) of the participants identified themselves as service users, 29% (9/31) identified as community members of individuals with mental health concerns, 26% (8/31) identified as peer support specialists, 26% (8/31) identified as a family member of a loved one with mental health or substance use challenges, and 26% (8/31) identified as nonprofit service workers. Other reported roles included recovery coaches, clinicians, administrators, other health service providers, and *other* for those who did not identify with any listed role (Table 1).

We identified 104 codes and 7 themes emerged from the data analysis. The themes identified were emotional holding and containment, training feedback, principles and practices of eCPR, implementation, connection in a digital environment, skills practice, and shared experiences.

Theme 1: Emotional Holding and Containment

Emotional holding and containment was the most prominent theme, with 25.9% (27/104) of codes related to this theme. Participants reported feeling accepted, heard, and seen. Through emotional holding, the participants reported that they recognized and understood the feelings of the people they were working with. Participants reported feeling held and being able to hold someone. A participant shared the following:

> *The expression that I resonate with is I felt held, you know, I felt emotionally held and when I am very*

> *anxious and fearful scared, being held and seen is super important.*

Another participant noted the following:

> *One of the things I realized that I needed to do more was just being able to be with somebody and not fill the space but let them be and just let them speak and do more listening.*

Another participant said the following:

> *This training was a reminder that it is okay to honor my feelings, that they are telling me something...If I am with someone else, I can feel what I feel and still be connected with that person in their feelings.*

The participants discussed the idea of being connected to other people's feelings and emotions. They reported that they were able to provide a safe container for others and felt emotionally contained during training.

Theme 2: Training Feedback or Training Format

Participants discussed their experiences with the eCPR training, especially compared with the other trainings they participated in. Training feedback was the second most predominant theme, with 24% (25/104) of codes related to this theme. Feedback from participants about the eCPR training experience varied, with both positive and negative feedback. Participants who liked the unscripted, unconstructed format of the eCPR training reported that the format allowed for them just to be witnesses and to be present in that moment. A participant shared the following comment:

> *It was a complete kind of like 180 degrees for me from everything else that I have learned as far as the mutuality, as far as the lived, shared, a lived experience and having that peer support...You know, you [are] just there, you [are] just there, however, it goes, you are just with it.*

Although several participants preferred the unconstructed and unscripted format, some participants suggested that concrete instructions about eCPR training and planned activities would have helped prepare them. A participant made the following comment:

> *Since some people like the more nebulous unconstructed freeform way of doing the training maybe for the sake of people like me, give us a little preparation like this is going to be very uncomfortable. You are not going to know what we are talking about. This training is a practice in how to listen. You are going to be listening without doing anything but being emotionally present.*

Participants suggested that mindfulness exercises be incorporated at the beginning of the training to prepare them for the emotional holding and containment experienced during training. A participant made the following comment:

> *The only thing that I would have loved to have seen, and perhaps it could even enhance the training- what I would have liked is that at the beginning of each training session somebody led a guided mindfulness practice.*

In addition, the participants suggested that training facilitators should play an active role and direct the training instead of being active participants in the training. A participant noted the following:

> *For me, it kind of felt a little free flow...I wanted someone responsible for us as a group. Someone to be the concrete space.*

The participants also suggested that the PowerPoint used for the training needed to be updated to align further with the content presented in the current eCPR workbook. A suggestion was made to include videos of eCPR in practice or encourage facilitators to role-play and model the skills during the training to provide further examples of practice.

Theme 3: Principles and Practices of eCPR

In all, 15.4% (16/104) of codes identified related to the principles and practices of eCPR. The participants understood eCPR as a framework for interacting with others. Participants noted that they found training empowering, as it provided freedom to practice with no expectations of fixing or finding a solution to someone's problems. The training's lack of concreteness and lack of a script on how to conduct the intervention was described as freeing. Participants found eCPR to be an easy framework that led to positive experiences and deeper connections. A participant shared the following comment:

> *What this emotional CPR does, it lets you respectfully experience emotions with someone. You let them have their moment and experience it and avoid assumptions and your opinion and what you want...We need this so bad here.*

The participants appreciated the lack of need to fix someone or to come up with a diagnosis. Participants noted that the lack of *the need to fix* led to a trusting and safe environment. A participant made the following comment:

> *eCPR allows us to move beyond the "what is wrong with you" or the stamp of a diagnosis. You say yeah! Of course, you would feel this way; it is hard to feel like this. And going back to that permission to own whatever emotions are coming up and be able to share those.*

In comparing the principles and practices of eCPR with those of other trainings the participants had previously taken, participants appreciated that they did not have to memorize a script or any other scripted skills. Rather than learning specific processes and technical content, the participants realized that they just needed to be present and listen deeply to promote recovery through eCPR. A participant shared the following:

> *Some of the things that we have learned are that you do not have to have the special words...I came away with a different level of confidence because when I was doing the motivational interviewing, I thought maybe I should get some cards or something so that*

> *I can do it...but now I do not have to memorize anything to do that* [eCPR]

Theme 4: Implementation

In all, 12.5% (13/104) of codes related to implementation were identified. The participants were excited about the training and raised issues related to the feasibility of eCPR implementation in their organization. The participants discussed the importance of obtaining stakeholder buy-in from management, mainly because of the paucity of research on the effectiveness of eCPR on specific populations. Participants noted that there was a need for organizational and personal investments. Organizational investment is needed because the training is costly, and training cannot be conducted in a short period. Staff should be removed from the clock to attend training. A participant made the following comment:

> *The thing that is challenging is that this is not information that you slap on a PowerPoint and show 30 at a time in 24-hour increments and then call it done...It takes time and to a certain extent involves internal work and kind of a level of reflection for everybody that's involved.*

The same participant also noted that staff participating in the training also had to be ready and willing to share some of their personal issues. This is because the training is structured with the use of real-life problems by the participants, which encourages people to share private information that they usually would not want to share with strangers. A participant made the following comment*:*

> *...and so when you sign up for this training program, you are signing up at a very personal level than anything else you have ever done.*

Participants were concerned that eCPR may not be appropriate for some of the clients they work with because individuals with certain disorders lack the ability to connect emotionally based on the nature of the disorder. A participant made the following comment:

> *I think a concern was it does not fit with all diagnoses. So, it will not necessarily work for everyone, especially when you get into your personality disorders like narcissistic personalities.*

The participants also raised concerns about workplace culture and how eCPR could shift how their organizations currently function. A participant commented as follows:

> *I think it would be a massive culture shift for our hospitals and how to do it [implement eCPR]...and how to get it rolled out that was one of the concerns is it is such a huge culture change.*

Implementation concerns related to personnel have also been raised. Some participant organizations were under the impression that they needed to have a peer support specialist on staff to implement eCPR. Participants suggested that agencies would need clear guidance related to a process for implementing eCPR, as most agencies had not yet developed a plan for how eCPR would be incorporated into their organization.

Theme 5: Connection in a Digital Environment

In all, 10.6% (11/104) of codes related to connecting in a virtual environment emerged. The participants were surprised by the connections they made despite training in a virtual environment. The participants reported feeling safe and trusting enough to open up to others. A participant made the following comment that was echoed by others:

> *The depth of communication and the sense of trust, even though we were on the computer and had never met these people before...There was a skill there on the human side that was very well done. I felt renewed, refreshed, comforted, safe. I felt very positive.*

The participants came into the training skeptical about attending virtual training but were surprised by the more profound sense of connection they had in training with not only people they were familiar with but also people they had never met. A participant commented as follows:

> *I was really impressed about how successful it could be done by Zoom. I was really coming in pretty skeptical, and when we did the real play where...everybody signed off on the video. It was just the two participants. You know, it really felt like you had that intimacy.*

Theme 6: Skills Practice

Experiential and scenario-based training, especially the *real plays*, was found to be effective in teaching eCPR. Real plays refer to role-playing exercises based on real-life situations that were part of the training. Of the 104 codes, 9 (8.6%) were identified related to skills practice and experiential learning. The participants found it easy to translate eCPR skills into practice compared with the other trainings they had attended. The participants reported using the skills they had learned during training with clients, peers, and friends. A participant commented on their first experience using their eCPR skills outside of the training environment:

> *I had situations literally within a week of our last session where a client came in, a husband and a wife who was just distraught...I just got certified. I went right into the eCPR mode, so I had a few incidents like that since then...*

Another participant shared the following*:*

> *I had an opportunity to use the eCPR with a co-worker who was expressing frustration, and actually on my part I was thinking, "I do not have to fix this, I just have to be present." You know, it was more of, I am empathizing with the position that she was in.*

Overall, participants found eCPR skills easy to incorporate into their everyday interaction with their clients, partly because eCPR used real-life client problems versus role-play with made-up scenarios as evidenced by this comment from a participant:

> *The "real plays" versus the role plays were the number one difference between any training because*

> *normally it is always a role play, you are given a character you kind of work through the skill set...It was helpful to actually...talk about something of value and not just a thing that you are kind of going through the motions. I think it really brought it home.*

Theme 7: Shared Experiences

The participants' shared experiences were identified as an emerging theme. Shared experiences are both those that individuals have in common and an individual confiding their experience with another individual or group. Of the 104 codes, 3 (2.8%) related to shared experiences arose. The participants reported that shared experiences led to a deep sense of connection despite being in different locations across the United States. A participant made the following comment:

> *Everyone [in the training group] was from across the country. So, it was a huge perspective, you know...Having that different perspective across the country actually was a bonding experience for us because it was a shared experience, even though we were not in the same local area going through the same thing.*

Through shared experiences, participants were able to connect to deeper emotional levels. A participant commented as follows:

> *I found myself wanting to interrupt and offer affirmations. However, I held back when I saw how much it meant to her to be able to express how she was feeling and her experience, and when I held back and the more that I listened, we both connected.*

Discussion

Principal Findings

This study aimed to examine the participants' experiences after attending the web-based eCPR training. The analysis of the focus group discussions regarding participants' attitudes toward and experiences with eCPR training highlighted seven main themes: (1) emotional holding and containment, (2) training feedback, (3) principles and practices of eCPR, (4) implementation, (5) connection in a digital environment, (6) skills practice, and (7) shared experiences.

Many participants appreciated the format of the eCPR training and the opportunity to engage in this process with hands-on experience. They expressed feelings of relief in the freedom to practice eCPR teachings and use the eCPR framework without having to *fix* or always be solution driven. The process of connecting in eCPR involves active listening, being present, and fostering a safe environment [5]. The participants found that when the focus was shifted from a solution-driven approach to being present with another person, they were most likely to be in the moment and be with the other person as they experienced their emotions. They appreciated the value of the idea that not everyone is looking for a solution to their experiences, and instead, what they can do is be an active listener while someone talks through and processes what they are experiencing at the moment.

Evidence indicates that participants primarily had a positive training experience. Participants found eCPR training to be a valuable resource for learning new skills when engaging with an individual who may be in distress or experiencing a mental health crisis. These skills included active listening, integrating shared experiences into practice, the ability to connect with others on an emotional level even on a web-based platform, and how to emotionally hold someone in their time of distress and need. Douglas [20] defines containment as "the ability to receive and understand another's emotional communication, process it, and communicate understanding and recognition back to the other person." Both holding and containment are essential for healing and recovery [21]. Emotional holding and containment led to participants feeling a deeper sense of trust as they were trusting strangers with personal stories. Not only did the participants report that they felt a sense of emotional holding and containment when participating in the training, but they could actually practice these techniques in their work. eCPR has shown promising evidence for increasing supportive behaviors toward individuals who experience mental health challenges and trauma [5] and can prove useful in both clinical and nonclinical environments such as hospitals, community centers, nonprofits, workplaces, and the community at large.

Multiple participants expressed concern over how the teachings of eCPR would be implemented within their communities—primarily workplaces. However, the concern was not about the actual practices of eCPR but rather about staffing needs and requesting clarification on whether peer support specialists need to be hired to do this work to be able to present to managerial staff. Although it is helpful to have a peer support specialist on the staff, it is not necessary for implementing eCPR. With regard to implementation, participants reported feeling personally invested in this initiative and generally felt that their workplace culture would be accepting of implementing the lessons they learned.

Future research can expand on the impacts reflected in this study by examining the impact of eCPR training on subsets of the population. In addition, research can examine the impact of eCPR training when implemented in various community settings (ie, schools, police stations, medical offices, and apartment complexes). Future studies can focus specifically on the impact of the training has on participants based on their age, mental health diagnosis, role, and race. In addition, future research can be directed to evaluate the effectiveness and impact of in-person eCPR training versus that of eCPR training delivered in a web-based environment. Finally, to answer the call of the United Nations, research could evaluate the impact that implementing eCPR has on community mental health as a whole (ie, levels of anxiety, depression, and distress; feelings of loneliness and isolation; and suicidality). Finally, future studies may attempt to collect more specific details regarding the unique roles and perspectives of the participants. For instance, a clinician who might also be a service user may have differing viewpoints depending on the lens they use. Examining these topics related to eCPR will provide evidence-based insights to inform the future implementation of eCPR training.

Limitations

This study has some limitations. First, this is the first study to evaluate participants' experiences in eCPR training. Although the sample size was sufficient to reach saturation, it was not sufficient to stratify the experiential data by demographic characteristics (role, race, and age). In addition, consistent with the peer-academic partnership [8], participants were not asked to disclose any mental health diagnoses; thus, we were unable to stratify the results on that factor either. Focus group participants were recruited as part of a convenience sample, which may have led to a bias in the results. In addition, the training and focus groups were conducted on the web; thus, the results may only be generalizable to web-based environments.

Conclusions

Although a previous study by the authors evaluated the feasibility and preliminary effectiveness of eCPR training, to the authors' knowledge, this is the first study to examine participants' experiences in attending the eCPR training—a community mental health education and training program developed by individuals with a history of mental health challenges and/or trauma, designed to be delivered by such individuals and other community members. In light of the COVID-19 pandemic, the increased or exacerbated mental health challenges individuals face owing to the pandemic and the increased amount of trauma due to COVID-19 fatalities, the United Nations recognizes an increased need for community-based support and wide-scale community mental health psychoeducation. The eCPR training program answers this call by providing a framework for engaging in community-based support without a clinical degree or formal environment. It has been proven to be feasible and effective in its delivery [5]. It is a practice that community members can use to help one another through difficult times and may enhance social connection in a time of disconnection and physical distancing.

Conflicts of Interest

KLF offers consulting services through Social Wellness, LLC. SC is a paid part-time employee of the National Empowerment Center.

Multimedia Appendix 1
Focus group interview guide.
[PDF File (Adobe PDF File), 51 KB - formative_v6i6e32219_app1.pdf]

References

1. Key substance use and mental health indicators in the United States: results from the 2019 National Survey on Drug Use and Health (HHS Publication No. PEP20-07-01-001). Substance Abuse and Mental Health Services Administration. Rockville, MD: Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration; 2020. URL: https://www.samhsa.gov/data/sites/default/files/reports/rpt29393/2019NSDUHFFRPDFWHTML/2019NSDUHFFR090120.htm [accessed 2021-01-04]
2. Barnett ML, Huskamp HA. Telemedicine for mental health in the United States: making progress, still a long way to go. Psychiatr Serv 2020 Feb 01;71(2):197-198 [FREE Full text] [doi: 10.1176/appi.ps.201900555] [Medline: 31847735]
3. Czeisler M, Lane RI, Petrosky E, Wiley JF, Christensen A, Njai R, et al. Mental health, substance use, and suicidal ideation during the COVID-19 pandemic - United States, June 24-30, 2020. MMWR Morb Mortal Wkly Rep 2020 Aug 14;69(32):1049-1057 [FREE Full text] [doi: 10.15585/mmwr.mm6932a1] [Medline: 32790653]
4. Policy brief: COVID-19 and the need for action on mental health. United Nations. 2020. URL: https://unsdg.un.org/sites/default/files/2020-05/UN-Policy-Brief-COVID-19-and-mental-health.pdf [accessed 2004-01-21]
5. Myers AL, Collins-Pisano C, Ferron JC, Fortuna KL. Feasibility and preliminary effectiveness of a peer-developed and virtually delivered community mental health training program (emotional CPR): pre-post study. J Particip Med 2021 Mar 04;13(1):e25867 [FREE Full text] [doi: 10.2196/25867] [Medline: 33661129]
6. SAMHSA's working definition of recovery: 10 guiding principles of recovery. Substance Abuse and Mental Health Services Administration. 2012. URL: https://store.samhsa.gov/sites/default/files/d7/priv/pep12-recdef.pdf [accessed 2004-01-21]
7. WHO QualityRights tool kit: assessing and improving quality and human rights in mental health and social care facilities. World Health Organization. 2012. URL: https://apps.who.int/iris/bitstream/handle/10665/70927/9789241548410_eng.Pdf?sequence=3 [accessed 2004-01-21]
8. Fortuna K, Barr P, Goldstein C, Walker R, Brewer L, Zagaria A, et al. Application of community-engaged research to inform the development and implementation of a peer-delivered mobile health intervention for adults with serious mental illness. J Particip Med 2019;11(1):e12380 [FREE Full text] [doi: 10.2196/12380] [Medline: 32095314]
9. Principles of community engagement: second edition. Centers for Disease Control and Prevention. 2011. URL: https://www.atsdr.cdc.gov/communityengagement/pdf/PCE_Report_508_FINAL.pdf [accessed 2004-01-21]
10. Almeida M, Day A, Smith B, Bianco C, Fortuna K. Actionable items to address challenges incorporating peer support specialists within an integrated mental health and substance use disorder system: co-designed qualitative study. J Particip Med 2020 Nov 26;12(4):e17053 [FREE Full text] [doi: 10.2196/17053] [Medline: 33242015]

11. Fortuna KL, Muralidharan A, Goldstein CM, Venegas M, Glass JE, Brooks JM. Certified peer specialists' perspective of the barriers and facilitators to mobile health engagement. J Technol Behav Sci 2020 Dec;5(4):318-323 [FREE Full text] [doi: 10.1007/s41347-020-00138-7] [Medline: 33163620]
12. Fortuna KL, Myers AL, Walsh D, Walker R, Mois G, Brooks JM. Strategies to increase peer support specialists' capacity to use digital technology in the era of COVID-19: pre-post study. JMIR Ment Health 2020 Jul 23;7(7):e20429 [FREE Full text] [doi: 10.2196/20429] [Medline: 32629424]
13. Fortuna KL, Naslund JA, LaCroix JM, Bianco CL, Brooks JM, Zisman-Ilani Y, et al. Digital peer support mental health interventions for people with a lived experience of a serious mental illness: systematic review. JMIR Ment Health 2020 Apr 03;7(4):e16460 [FREE Full text] [doi: 10.2196/16460] [Medline: 32243256]
14. Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 2018;52(4):1893-1907 [FREE Full text] [doi: 10.1007/s11135-017-0574-8] [Medline: 29937585]
15. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 2013 Nov 27;310(20):2191-2194. [doi: 10.1001/jama.2013.281053] [Medline: 24141714]
16. IBM SPSS statistics for mac. IBM Corporation. 2020. URL: https://www.ibm.com/products/spss-statistics [accessed 2004-01-21]
17. Watkins DC. Rapid and rigorous qualitative data analysis: the "RADaR" technique for applied research. Int J Qual Methods 2017 Jun 08;16(1):160940691771213 [FREE Full text] [doi: 10.1177/1609406917712131]
18. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
19. Candela AG. Exploring the function of member checking. Qual Rep 2019 Mar 24;24(3):619-628 [FREE Full text] [doi: 10.46743/2160-3715/2019.3726]
20. Douglas H. Containment and reciprocity: integrating psychoanalytic theory and child development research for work with children. Florence, Italy: Routledge; 2007:24-44.
21. Blackwell D. Holding, containing and bearing witness: the problem of helpfulness in encounters with torture survivors. J Soc Work Pract 1997 Nov;11(2):81-89. [doi: 10.1080/02650539708415116]

Abbreviations

eCPR: Emotional Connecting, Empowering, Revitalizing
PI: principal investigator

Edited by J Torous; submitted 19.07.21; peer-reviewed by K Machin, D Smit; comments to author 18.09.21; revised version received 04.01.22; accepted 10.02.22; published 30.06.22.

Please cite as:
Myers AL, Mbao M, Kadakia A, Collings S, Fortuna KL
Experiences of Community Members Engaged in eCPR (Emotional Connecting, Empowering, Revitalizing) Training: Qualitative Focus Group Study
JMIR Form Res 2022;6(6):e32219
URL: https://formative.jmir.org/2022/6/e32219
doi:10.2196/32219
PMID:35771610

Original Paper

Therapeutic Alliance in Web-Based Treatment for Eating Disorders: Secondary Analysis of a Randomized Controlled Trial

Claudia Stoeten[1,2], MSc; Hein Arnoud de Haan[1,3], MD, PhD; Marloes Gerda Postel[4], PhD; Marjolein Brusse-Keizer[5], PhD; Elke Daniëlle ter Huurne[4], PhD

[1]Tactus Addiction Care, Deventer, Netherlands

[2]Mediant, Hengelo, Netherlands

[3]Nijmegen Institute for Scientist-Practitioners in Addiction (NISPA), Nijmegen, Netherlands

[4]Saxion University of Applied Sciences, Enschede, Netherlands

[5]Medical School Twente, Medisch Spectrum Twente, Enschede, Netherlands

Corresponding Author:
Hein Arnoud de Haan, MD, PhD
Tactus Addiction Care
Keulenstraat 3
7418 ET
Deventer, 7400 AD
Netherlands
Phone: 31 883822887
Fax: 31 570500115
Email: h.dehaan@tactus.nl

Abstract

Background: In face-to-face therapy for eating disorders, therapeutic alliance (TA) is an important predictor of symptom reduction and treatment completion. To date, however, little is known about TA during web-based cognitive behavioral therapy (web-CBT) and its association with symptom reduction, treatment completion, and the perspectives of patients versus therapists.

Objective: This study aimed to investigate TA ratings measured at interim and after treatment, separately for patients and therapists; the degree of agreement between therapists and patients (treatment completers and noncompleters) for TA ratings; and associations between patient and therapist TA ratings and both eating disorder pathology and treatment completion.

Methods: A secondary analysis was performed on randomized controlled trial data of a web-CBT intervention for eating disorders. Participants were 170 females with bulimia nervosa (n=33), binge eating disorder (n=68), or eating disorder not otherwise specified (n=69); the mean age was 39.6 (SD 11.5) years. TA was operationalized using the Helping Alliance Questionnaire (HAQ). Paired t tests were conducted to assess the change in TA from interim to after treatment. Intraclass correlations were calculated to determine cross-informant agreement with regard to HAQ scores between patients and therapists. A total of 2 stepwise regressive procedures (at interim and after treatment) were used to examine which HAQ scores predicted eating disorder pathology and therapy completion.

Results: For treatment completers (128/170, 75.3%), the HAQ-total scores and HAQ-Helpfulness scores for both patients and therapists improved significantly from interim to post treatment. For noncompleters (42/170, 24.7%), all HAQ scores decreased significantly. For all HAQ scales, the agreement between patients and therapists was poor. However, the agreement was slightly better after treatment than at interim. Higher patient scores on the helpfulness subscale of the HAQ at interim and after treatment were associated with less eating disorder psychopathology. A positive association was found between the HAQ-total patient scores at interim and treatment completion. Finally, posttreatment HAQ-total patient scores and posttreatment HAQ-Helpfulness scores of therapists were positively associated with treatment completion.

Conclusions: Our study showed that TA in web-CBT is predictive of eating disorder pathology and treatment completion. Of particular importance is patients' confidence in their abilities as measured with the HAQ-Helpfulness subscale when predicting posttreatment eating disorder pathology and treatment completion.

(JMIR Form Res 2022;6(6):e33813) doi:10.2196/33813

KEYWORDS

therapeutic alliance; TA; treatment completion; cognitive behavioral therapy; CBT; web-CBT; eating disorders

Introduction

Background

Eating disorders (EDs) are related to serious physical, psychological, and social consequences and are characterized by a chronic character and high treatment costs [1,2]. However, many patients have EDs for years before receiving treatment [3,4]. Access to face-to-face treatment of ED is often limited because of personal barriers, such as feelings of shame and fear of stigmatization, and intervention-related barriers, such as costs, geographic distance, and lack of availability [5-9]. Web-based alternatives, which may encompass website- and mobile app–based treatment programs, might show promise. Web-based treatment was shown to be effective in reducing ED psychopathology [10-22], and it can improve access to ED treatment compared with face-to-face treatment [15,17,19]. Web-based treatment provides the added advantages of approachability, relative anonymity, and widespread 24-hour access, which are considered important benefits for patients with ED [15].

One particularly important facet of face-to-face treatment is the therapeutic alliance (TA) between therapists and patients [23,24]. Although there are various ways to define the concept of TA [23], all definitions have in common that TA can best be characterized by the degree of agreement between a therapist and a patient concerning the goals and tasks of the treatment and suggest the presence of an affective bond [23-26].

TA was shown to be predictive of treatment completion and outcomes in the general population [27]. The quality of TA was also shown to be predictive of treatment completion and outcomes in face-to-face ED treatment [28-30]. However, the predictive value of TA for treatment outcomes in patients with ED varies between studies and between patient groups [28]. More specifically, the predictive value of TA for treatment outcome is less obvious for patients with bulimia nervosa (BN) than for patients with anorexia nervosa [28]. Overall, the predictive value of TA for treatment outcomes is associated with small to medium effect sizes [30].

With regard to web-based treatment in the general population, multiple studies have demonstrated that the strength of the TA during treatment can be improved without face-to-face contact with a therapist [31-37]. However, compared with face-to-face treatment, much less is known about important predictors in the development of TA in the context of web-based treatment [31,38]. Studies focusing on TA in web-based treatment are often methodologically inferior to those focusing on face-to-face treatment [31,38].

Few studies have been conducted on the role of TA with regard to treatment outcomes and adherence in web-based treatment for ED [11,38]. It was found that TA was rated high over the course of ED treatment [39,40]. It was also found that higher TA ratings were associated with better treatment outcomes [25,27,40]. Furthermore, some evidence indicates that the extent of TA during web-based treatment of ED is positively associated with treatment adherence [11,29,41].

Objectives

Concerning the effects of TA on treatment effectiveness, it is also important to emphasize the degree of agreement between the therapist and patient perspectives [42,43]. In the general population, it has been shown that convergent patient-therapist ratings over the course of treatment predict a better treatment outcome [42,43]. It was also found that, for face-to-face treatment in the general population, therapist ratings of the TA are not as predictive of treatment outcomes as the TA ratings provided by patients [43].

This study focused on web-based cognitive behavioral therapy (web-CBT) for female patients with ED and aimed to investigate (1) TA ratings measured at interim and after treatment, separately for patients (treatment completers and noncompleters) and therapists; (2) the degree of agreement between therapists and patients (treatment completers and noncompleters) for TA ratings; and (3) associations between patient and therapist TA ratings and both ED pathology and therapy completion. We hypothesized that the TA would increase from interim to post treatment for both patients and therapists and that there would be stronger agreement between therapists and patients who completed treatment than between therapists and patients who did not complete treatment. Furthermore, we hypothesized that TA ratings provided by patients and therapists are predictors of eating disorder pathology, particularly after treatment. Finally, we hypothesized that TA ratings would be positively associated with treatment completion.

Methods

Study Design

A secondary analysis was conducted on the data from a randomized controlled trial (RCT) investigating a web-CBT intervention for EDs. The study design, procedures, and results of the RCT are described in detail elsewhere [19,41,44,45]. Recruitment for the RCT was conducted from March 2011 to December 2013. Information on the study was disseminated through announcements on ED-related websites, forums, and newspaper advertisements.

Ethics Approval

All participants provided written informed consent, and the study was approved by the Medical Ethics Committee of the Medical Spectrum Twente (NL31717.044.010, P10-31) and registered in the Netherlands Trial Registry (NTR2415).

An RCT [19,44] compared a web-CBT intervention group to a waiting list control group. Participants were stratified by ED type (BN, binge eating disorder [BED], or eating disorder not otherwise specified [EDNOS]). Participants in the intervention group started web-CBT immediately, whereas those in the control group had to wait 15 weeks after randomization. Both completers and noncompleters completed the questionnaire.

Outcomes were measured before, during, and after web-CBT and at 3-, 6-, and 12-month follow-up.

For the current analysis, measurements at the interim (after the first part of treatment) and after treatment were used. As this study did not focus on the efficacy of the treatment but on the interim and posttreatment measurements of the TA, the data from the intervention phase of the study of both the initial intervention and control groups were merged.

Participants

The participants of this study were female patients with a Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) diagnosis of BN, BED, or EDNOS who completed the first part of the web-based CBT and completed the interim questionnaire. In addition to the DSM-IV classification, the inclusion criteria for the RCT were (1) age ≥18 years, (2) access to the internet, (3) fluent in Dutch, (4) referral from a general practitioner, and (5) to be within 85% of the target weight established by the table of height and weight limits of MINI-Plus [46,47]. Exclusion criteria were as follows: (1) suicidal ideation, (2) receiving psychological or pharmaceutical treatment for any ED within the past 6 months, (3) pregnancy, and (4) expected absence of 4 weeks or longer during the treatment period of 15 weeks.

Intervention

The web-based treatment program, Etendebaas (English translation: “Look at your eating”), included a structured 15-week web-CBT that was designed within a secure web-based application [19,44]. The treatment program consisted of 2 parts and included 16 treatment modules, with at least 21 scheduled asynchronous contact moments and 10 homework assignments. The first part aimed to analyze participants’ ED attitudes and behaviors, whereas the second part focused on behavioral changes. All treatment modules were completed by the patients in a fixed order, and it was not possible to skip a module.

CBT [48-50] and motivational interviewing [51,52] were the fundamental elements of the intervention, which included techniques such as psychoeducation, self-monitoring through daily diary entries, thought restructuring, problem-solving, and relapse prevention. In their personal files, patients could read and respond to the therapist’s messages and complete homework assignments. The treatment protocol prescribed regular contact between patients and their therapists, with therapists responding to the patients’ messages and assignments within 3 working days.

A total of 17 therapists carried out web-based treatments, including 2 male therapists and 15 female therapists. Therapists had either a bachelor’s degree in nursing or social work or a master’s degree in psychology and received specific training for web-based treatment. A comprehensive manual was available, which included a detailed description of all treatment modules and safety protocols. The treatment also included web-based coaches and support from a multidisciplinary team (psychologists, psychotherapists, addiction medicine physicians, psychiatrists, and dieticians) who were available for consultation. The participating therapists did not have knowledge of the TA scores of the patients and did not receive any instructions regarding investing in improving the TA. However, within the regular web-based treatment protocol, the core task of a therapist is to build and maintain the TA.

Measures

Therapeutic Alliance

TA was measured using the Dutch patient and therapist version of the Helping Alliance Questionnaire (HAQ) [53,54]. The HAQ is a self-report questionnaire that measures the strength of a therapeutic patient-therapist alliance [55]. The therapist’s version was derived from the patient’s version and was compatible. The Dutch version of the HAQ has 11 items scored on a 5-point Likert scale (1=totally disagree, 2=disagree, 3=neutral, 4=agree, and 5=totally agree) [54]. The HAQ contains two subscales: (1) cooperation (5 items), reflecting the perception of the patient on the cooperation with a care provider or vice versa, and (2) helpfulness (6 items), which reflects a patient’s or therapist’s confidence in their own capacity to improve the situation. The HAQ-total score was determined as the sum of the subscale scores. This study found the patient version of the HAQ to be internally reliable at the interim measurement: Cronbach α of .81 for the cooperation subscale, Cronbach α of .81 for the helpfulness subscale, and Cronbach α of .87 for the total HAQ score. The therapist version of the HAQ was also internally reliable: Cronbach α of .77 for the cooperation subscale, Cronbach α of .78 for the helpfulness subscale, and Cronbach α of .87 for the total HAQ score.

Eating Disorder Psychopathology

Changes in the clinical severity of ED psychopathology were measured using the total score of the Eating Disorder Examination Questionnaire (EDE-Q) [56]. The EDE-Q is a widely used validated self-report scale based on Eating Disorder Examination interviews. The instrument focuses on the previous 28 days to assess important behavioral and attitude aspects of ED and the severity of ED psychopathology. The EDE-Q consists of 36 items, with four subscales (restraint, eating concern, shape concern, and weight concern). Items are scored on a 7-point Likert scale (range 0=not one single day–6=every day), with a higher score reflecting more psychopathology. Subscale scores were obtained by averaging the items for each subscale, whereas the total EDE-Q score was obtained by summing the subscale scores. Previous research indicates that the EDE-Q demonstrates acceptable internal consistency (Cronbach α ranging from .77 to .84) [57-59].

Treatment Completion

Participants were considered completers when they (1) had attended all 16 treatment modules with at least 21 contact moments with their personal therapist, (2) completed all 10 homework assignments, and (3) completed the at-interim and posttreatment questionnaires. Participants who stopped the treatment program before the completion of all treatment modules and completed the at-interim and posttreatment questionnaires were considered noncompleters. Therefore, treatment completion was operationalized using a dichotomous measure (yes or no).

Statistical Analysis

All analyses were conducted using SPSS for Windows (version 21; IBM Corp) [60]. Continuous variables were summarized using the mean with the associated SD or as the median with the associated IQR for normally and nonnormally distributed data, respectively. Categorical variables were summarized as frequencies with corresponding percentages. Sum scores were computed for the HAQ-total score and separately for the cooperation and helpfulness subscales, both at interim and after treatment, and separately for patients and therapists.

Differences in demographic characteristics between completers and noncompleters were analyzed using independent 2-tailed *t* tests for continuous normally distributed data and Wilcoxon rank-sum tests for continuous nonnormally distributed data. Differences in categorical variables were analyzed using chi-square or Fisher exact tests (as appropriate).

Paired *t* tests were conducted to assess the change in TA ratings from the interim to the end of treatment for both therapists and patients. In the analyses, we stratified for completers and noncompleters because we expected an opposite pattern in TA ratings from interim to post treatment. Cohen $d=(\mu_1-\mu_2)/\sigma_{1,2}$ was calculated to determine the effect sizes for significant findings. Cohen defines *d* scores of 0.2, 0.5, and 0.8 as small, medium, and large effects, respectively [61]. An intraclass correlation coefficient (ICC) analysis was conducted to determine cross-informant agreement between patients (separately for completers and noncompleters) and therapists in TA ratings, both at interim and post treatment. To determine the strength of agreement, the guidelines drafted by Koo and Li [62] were used (<0.50: poor, between 0.50 and 0.75: moderate, between 0.75 and 0.90: good, and >0.90: excellent) [62].

Next, we examined whether the HAQ scores for patients and therapists were related to ED pathology by creating two linear regression models: one at interim and one post treatment. To do this, TA ratings of therapists and patients (completers and noncompleters combined) were first analyzed separately in univariate linear regression models. We merged the data of completers and noncompleters. The choice to combine completers and noncompleters was based on the following arguments: (1) this would increase statistical power; (2) this would provide fairer insights, as TA ratings of completers were expected to be overly positive specifically because these patients completed the treatment, whereas TA ratings of noncompleters may have been more critical; and (3) by including both groups, the range of ED pathology included in the analyses was broader, increasing the ecological validity of the results. Rating scores that were sufficiently related ($P \leq .15$) in these univariate analyses were entered into a multiple linear regression model. Owing to multicollinearity between the TA subscales and total scale within the patient or therapist groups, we entered the total scale or subscales with the highest explained variance (R^2) into the multiple linear regression model. Nonsignificant variables were removed individually until the explained variance deteriorated significantly.

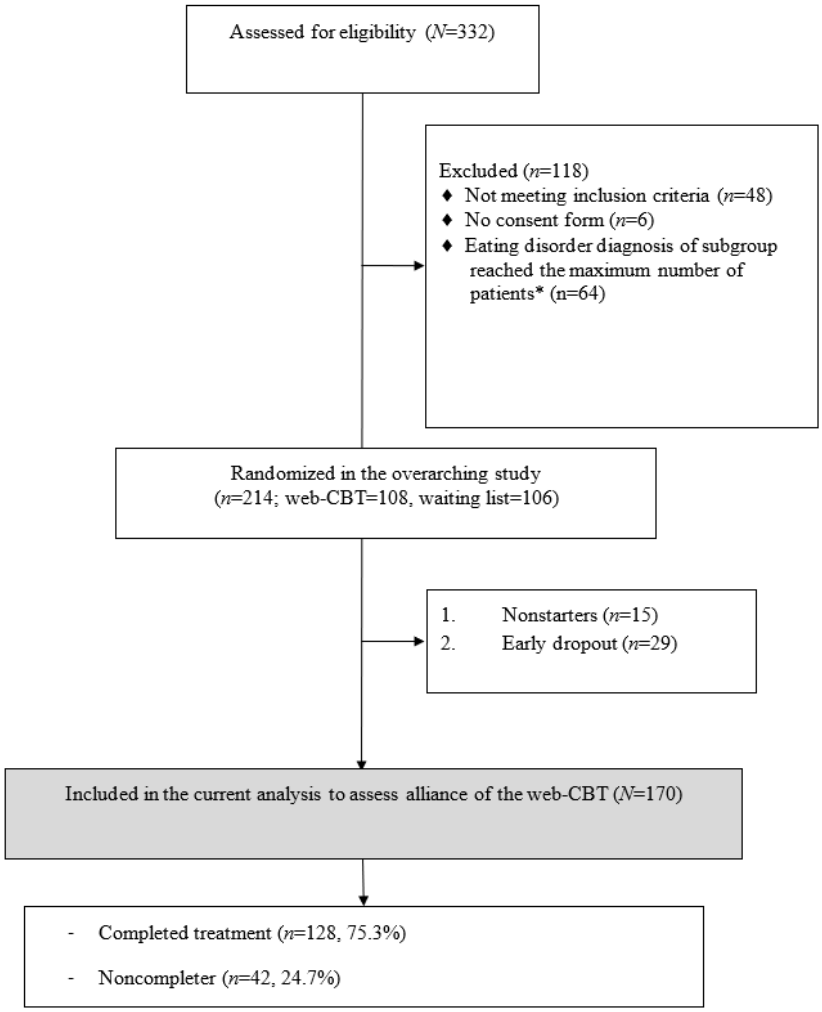
To assess whether TA ratings were related to treatment completion, we constructed two logistic regression models: one at interim and one post treatment. These models were constructed identically to the construction of the previously described multiple regression models. Owing to multicollinearity between the TA subscales and the total scale for the patient or therapist groups at each time point, we entered the scale(s) that produced the best model fit (−2 log likelihood). Nonsignificant variables were removed one by one until the −2 log likelihood deteriorated significantly. Nagelkerke R^2 was used to estimate the pseudoproportion of the variance. Two-sided significance levels were set to 0.05 in all measurements.

Results

Inclusion Process

Figure 1 presents a flowchart of the inclusion process used in this study. A total of 214 participants were included in an earlier RCT [19,44]. Of these, 128 (59.8%) completed the treatment. Of the participants who did not complete the web-based CBT (86/214, 40.1%), 15 never started treatment (nonstarters), and 29 stopped the treatment before the end of the first part of the treatment (early dropouts). The 44 participants did not complete the interim questionnaire, so no information about their experiences with the TA was available. Therefore, these were excluded from the analysis. Of the 170 participants who were included in this study, 42 stopped the program during the second part of the web-based CBT. These participants were considered late dropouts and filled out the interim and posttreatment questionnaires, including the TA. In this study, late dropouts were defined as noncompleters, although the overall number of noncompleters in the RCT was higher (n=106), as it also included 15 nonstarters and 29 early dropouts.

Figure 1. Flowchart of the inclusion process of this study. *In the underlying RCT, power analysis was used to determine how many participants could be assigned to each subgroup. The number of patients included in the binge eating disorder (BED) and eating disorder not otherwise specified (EDNOS) subgroups reached the necessary number of patients that should be recruited within the subgroup based on the sample size calculation, and the necessary number of patients for the bulimia nervosa (BN) group was not reached [19,44]. web-CBT: web-based cognitive behavioral therapy.



Participants

The participant characteristics are reported in Table 1. The sample included 170 women with BN (n=33), BED (n=68), or EDNOS (n=69), and a mean age of 39.6 (SD 11.5) years. Completers reported significantly higher BMI scores than noncompleters. When stratifying for BMI categories (underweight, normal weight, and overweight), no significant differences were found between these groups.

Table 1. Participants characteristics for completers and noncompleters.

Variable	Overall (N=170)	Completers (n=128)	Noncompleters (n=42)	*P* value
Age (years), mean (SD)	39.6 (11.5)	40.5 (11.1)	36.6 (12.4)	.06
BMI (kg/m^2), mean (SD)	31.9 (6.9)	32.6 (6.9)	30 (6.8)	.04[a]
BMI (kg/m^2), n (%)				.25
Underweight: <18.5	6 (3.5)	3 (2.3)	3 (7.1)	
Normal weight: 18.5-25	21 (12.4)	15 (11.7)	6 (14.3)	
Overweight: >25	143 (84.1)	110 (85.9)	33 (78.6)	
Eating disorder, n (%)				.67
Bulimia nervosa	33 (19.4)	23 (18)	10 (23.8)	
Binge eating disorder	68 (40)	53 (41.4)	15 (35.7)	
Eating disorder not otherwise specified	69 (40.6)	52 (40.6)	17 (40.5)	
Living situation, n (%)				.31
Alone	38 (22.4)	31 (24.2)	7 (16.7)	
With others	132 (77.6)	97 (75.8)	35 (83.3)	
Level of education, n (%)				.70
Low	18 (10.6)	15 (11.7)	3 (7.1)	
Intermediate	59 (34.7)	44 (34.4)	15 (35.7)	
High	93 (54.7)	69 (53.9)	24 (57.1)	
Employment, n (%)				.28
Paid job	139 (81.8)	107 (83.6)	32 (76.2)	
No paid job	31 (18.2)	21 (16.4)	10 (23.8)	
Duration of eating disorder (years), n (%)				.13
1-5	24 (14.1)	14 (10.9)	10 (23.8)	
6-10	25 (14.7)	19 (14.8)	6 (14.3)	
11-20	53 (31.2)	39 (30.5)	14 (33.3)	
>20	68 (40)	56 (43.8)	12 (28.6)	
Professional treatment of eating disorder, n (%)				.48
Yes	77 (45.3)	56 (43.8)	21 (50)	
No	93 (54.7)	72 (56.3)	21 (50)	
Professional treatment, n (%)				.90
Yes	112 (65.9)	84 (65.6)	28 (66.7)	
No	58 (34.1)	44 (34.4)	14 (33.3)	
Medication use, n (%)				.09
Yes	77 (60.2)	19 (45.2)	96 (56.5)	
No	51 (39.8)	23 (54.8)	74 (43.5)	
Smoking, n (%)				.66
Yes	21 (12.4)	15 (11.7)	6 (14.3)	
No	149 (87.6)	113 (88.3)	36 (85.7)	
Alcohol use (experienced problematic), n (%)				.99
Yes	6 (4.2)	5 (4.6)	1 (2.9)	
No	136 (95.8)	103 (95.4)	33 (97.1)	
Drug use, n (%)				.26

Variable	Overall (N=170)	Completers (n=128)	Noncompleters (n=42)	*P* value
Yes	4 (2.4)	2 (1.6)	2 (4.8)	
No	166 (97.6)	126 (98.4)	40 (95.2)	
Gambling, n (%)				.99
Yes	4 (2.4)	3 (2.3)	1 (2.4)	
No	166 (97.6)	125 (97.7)	41 (97.6)	

[a]*P*<.05.

TA Before and After Treatment

In Table 2, TA ratings at interim and post treatment and the difference scores between the measurements are reported separately for completers and noncompleters and from both the patient and therapist perspectives. For completers, the HAQ-total and HAQ-Helpfulness scores improved significantly from interim to the end of treatment, with effect sizes ranging from small to medium. For noncompleters, all 3 types of HAQ scores significantly decreased, with medium to large effect sizes.

Table 2. Helping Alliance Questionnaire scores of patients and therapists at interim and posttreatment and difference scores.

	Interim scores		Posttreatment scores		Difference scores[a]			
	N	Value, mean (SD; range)	Value, N	Value, mean (SD; range)	Value, N	Value, mean (SD)	*P* value	Cohen *d*
Patients								
Completers								
HAQ-Co[b]	128	20.6 (2.4; 13.0-25.0)	126	20.8 (2.6; 15.0-25.0)	126	0.2 (2.1)	.25	0.10
HAQ-HE[c]	128	23.3 (3.2; 14.0-30.0)	126	24.7 (3.1; 16.0-30.0)	126	1.4 (3.1)	<.001[d]	0.45
HAQ-T[e]	128	43.9 (5.0; 30.0-55.0)	126	45.5 (5.1; 32.0-55.0)	126	1.6 (4.6)	<.001[d]	0.35
Noncompleters								
HAQ-Co	42	18.8 (3.0; 9.0-25.0)	31	16.0 (4.9; 5.0-23.0)	31	−3.1 (4.0)	<.001[d]	−0.78
HAQ-HE	42	20.4 (3.8; 11.0-29.0)	31	17.8 (5.1; 8.0-28.0)	31	−2.2 (3.6)	.002[d]	−0.60
HAQ-T	42	39.2 (5.9; 24.0-53.0)	31	33.9 (8.2; 13.0-46.0)	31	−5.3 (6.0)	<.001[d]	−0.89
Therapists								
Completers								
HAQ-Co	125	19.8 (1.9; 15.0-25.0)	126	20.0 (2.4; 12.0-25.0)	123	0.2 (2.2)	.42	0.07
HAQ-HE	125	22.6 (2.9; 11.0-28.0)	126	24.1 (3.0; 14.0-30.0)	123	1.5 (3.0)	<.001[d]	.50
HAQ-T	125	42.4 (4.4; 28.0-53.0)	126	44.1 (5.1; 27.0-55.0)	123	1.6 (4.7)	<.001[d]	0.35
Noncompleters								
HAQ-Co	42	19.3 (2.9; 13.0-25.0)	40	16.6 (3.3; 10.0-24.0)	40	−2.6 (2.8)	<.001[d]	−0.91
HAQ-HE	42	21.4 (3.3; 16.0-27.0)	40	18.6 (4.2; 10.0-29.0)	40	−2.6 (3.2)	<.001[d]	−0.82
HAQ-T	42	40.7 (5.9; 30.0-51.0)	40	35.2 (7.0; 22.0-53.0)	40	−5.2 (5.2)	<.001[d]	−0.99

[a]Difference score=posttreatment score−interim score.

[b]HAQ-CO: Helping Alliance Questionnaire Cooperation.

[c]HAQ-HE: Helping Alliance Questionnaire Helpfulness.

[d]*P*<.05.

[e]HAQ-T: Helping Alliance Questionnaire total.

Cross-Informant Agreement Between Patients and Therapists

The ICCs that represent agreement between therapists and patients regarding the TA (represented by the HAQ-total score and HAQ subscale scores) are presented in Table 3. Agreement between therapists and patients increased as treatment progressed. However, in general, agreement between patients and therapists was poor for both noncompleters and completers both at interim and post treatment.

Table 3. Intraclass correlations between therapists and patients.

	Value, *N* valid	Intraclass correlation	*P* value
HAQ[a]-Cooperation at interim[b]			
Completers	125	0.09	.17
Noncompleters	42	−0.16	.85
HAQ-Helpfulness at interim			
Completers	125	0.26	.002[c]
Noncompleters	42	−0.07	.67
HAQ-total at interim			
Completers	125	0.20	.014[c]
Noncompleters	42	−0.14	.82
HAQ-Cooperation post treatment[d]			
Completers	124	0.29	.001[c]
Noncompleters	30	0.41	.011[c]
HAQ-Helpfulness post treatment			
Completers	124	0.41	<.001[c]
Noncompleters	30	0.34	.029[c]
HAQ-total post treatment			
Completers	124	0.39	<.001[c]
Noncompleters	30	0.48	.003[b]

[a]HAQ: Helping Alliance Questionnaire.

[b]At interim, there were 3 missing participant scores.

[c]*P*<.05.

[d]Post treatment, there were 4 missing scores for completers and 12 for noncompleters.

Associations With Treatment Outcome

For ED pathology measured with the EDE-Q, Table 4 shows the results of the univariate regression analyses at the interim. All patients' HAQ scores at interim were found to be univariately negatively associated with the extent of ED psychopathology, as was the therapists' HAQ-Helpfulness score at interim. The subscales showed the best explained variance; therefore, we entered these subscales and not the total HAQ scale in the initial multiple regression model. After entering the patients' subscale scores and therapists' scores for the helpfulness subscale and completing the stepwise regression procedure, only the HAQ-Helpfulness score at the interim of patients was significantly negatively associated with ED pathology. The final interim model with the HAQ-Helpfulness scores of patients as the sole predictor explained 9.8% (F_1=17.03; *P*<.001) of the variance in eating disorder pathology.

As presented in Table 5, posttreatment HAQ scores of all patients were found to be univariately negatively associated with posttreatment ED psychopathology, as well as therapists' HAQ-Helpfulness and HAQ-total scores after treatment. Owing to multicollinearity and the best model of fit for the subscales, all patients' subscale scores and the therapists' helpfulness subscale scores were entered in the initial multiple regression model. After completing the stepwise regression procedure, the HAQ-Helpfulness score of the patients remained the only predictor that was negatively associated with ED pathology after treatment. After treatment, the HAQ-Helpfulness score of patients explained 22.3% (F_1=43.58; *P*<.001) of the variance in ED pathology.

Table 4. Univariate regression models for the at-interim association between the strength of the therapeutic alliance (Helping Alliance Questionnaire scores) and eating disorder pathology (Eating Disorder Examination Questionnaire), separately per subscale and as total score, and separately for patients and therapists.

Outcome variable at interim	Univariate coefficients (95% CI)	*P* value
Eating Disorder Examination Questionnaire		
Patients		
HAQ[a]-Cooperation	−0.08 (−0.14 to −0.01)	.02[b]
HAQ-Helpfulness	−0.10 (−0.15 to −0.05)	<.001[b]
HAQ-total	−0.06 (−0.09 to −0.03)	<.001[b]
Therapists		
HAQ-Cooperation	−0.04 (−0.12 to 0.05)	.40
HAQ-Helpfulness	−0.06 (−0.12 to 0.001)	.06
HAQ-total	−0.03 (−0.07 to 0.01)	.12

[a]HAQ: Helping Alliance Questionnaire.
[b]$P<.05$.

Table 5. Univariate regression models for the posttreatment association between the strength of the therapeutic alliance (Helping Alliance Questionnaire scores) and eating disorder pathology (Eating Disorder Examination Questionnaire), separately per subscale and as a total score, and separately for patients and therapists.

Outcome variable post treatment	Univariate coefficients (95% CI)	*P* value
Eating Disorder Examination Questionnaire		
Patients		
HAQ[a]-Cooperation	−0.09 (−0.13 to −0.04)	.001[b]
HAQ-Helpfulness	−0.12 (−0.16 to −0.08)	<.001[b]
HAQ-total	−0.07 (−0.09 to −0.04)	<.001[b]
Therapists		
HAQ-Cooperation	−0.03 (−0.10 to 0.04)	.39
HAQ-Helpfulness	−0.08 (−0.12 to −0.03)	.002[b]
HAQ-total	−0.03 (−0.06 to −0.01)	.02[b]

[a]HAQ: Helping Alliance Questionnaire.
[b]$P<.05$.

Associations With Treatment Completion

Table 6 shows the results of the univariate logistic regression analyses at the interim of the association between HAQ scores and treatment completion. All patients' HAQ scores at the interim measurement and the therapists' HAQ-Helpfulness and HAQ-total score at the interim were found to be univariately associated with treatment completion. We entered patients' HAQ-total scores and therapists' subscale scores into the initial multiple regression model because these resulted in the best model fit. In the final multivariate model, the HAQ-total score of patients at the interim measurement remained the only significant predictor of treatment completion, explaining 18.8% (−2 log likelihood=167.10; Nagelkerke R^2=0.188) of the pseudovariance in treatment completion.

Table 7 shows the results of the posttreatment univariate logistic regression analyses, focusing on HAQ scores and treatment completion. All the HAQ scores were found to be univariately positively associated with treatment completion. The HAQ-total scores of patients and the therapists' subscales were entered into the initial model because these resulted in the best model fit. In the final multiple regression model, both the HAQ-total scores of patients (odds ratio [OR] 0.30, 95% CI 1.18-1.55; $P\leq.001$) and HAQ-Helpfulness scores of therapists (OR 0.13, 95% CI 0.97-1.34; P=.12) were positively associated with treatment completion, explaining 59% (−2 likelihood=80.24; R^2=0.59) of the pseudovariance in treatment completion.

Table 6. Univariate regression models at interim Helping Alliance Questionnaire scores and treatment completion.

	Odds ratio (95% CI)	*P* value
Patients		
HAQ[a]-Cooperation	1.30 (1.12-1.51)	<.001[b]
HAQ-Helpfulness	1.29 (1.15-1.46)	<.001[b]
HAQ-total	1.18 (1.10-1.28)	<.001[b]
Therapists		
HAQ-Cooperation	1.12 (0.95-1.32)	.17
HAQ-Helpfulness	1.13 (1.01-1.27)	.03[b]
HAQ-total	1.07 (1.00-1.15)	.05[b]

[a]HAQ: Helping Alliance Questionnaire.
[b]*P*<.05.

Table 7. Univariate regression models posttreatment Helping Alliance Questionnaire scores and treatment completion.

	Odds ratio (95% CI)	*P* value
Patients		
HAQ[a]-Cooperation	1.58 (1.30-1.92)	<.001[b]
HAQ-Helpfulness	1.58 (1.34-1.87)	<.001[b]
HAQ-total	1.40 (1.23-1.58)	<.001[b]
Therapists		
HAQ-Cooperation	1.54 (1.32-1.80)	<.001[b]
HAQ-Helpfulness	1.51 (1.32-1.73)	<.001[b]
HAQ-total	1.27 (1.17-1.37)	<.001[b]

[a]HAQ: Helping Alliance Questionnaire.
[b]*P*<.05.

Discussion

Principal Findings

First, in line with our expectations, it is possible to examine and detect changes in the TA of web-based CBT. Our study showed that the HAQ-total and HAQ-Helpfulness scores for completers significantly increased from interim to post treatment, whereas for noncompleters, all 3 HAQ scores significantly decreased. This shows that, in general, patients who completed treatment experienced a TA that grew stronger over time, whereas those who did not complete treatment experienced a weaker TA that decreased over time. These findings were observed for both patients and therapists and confirmed the results of previous studies [44,63].

In addition, we found that although the ICCs of agreement between patients and therapists increased from interim to posttreatment measurement, the overall agreement about the degree of TA remained relatively poor. This might partly be because of differences in perceptions between patients and therapists regarding what the TA entails. Patients are more concerned with the helpfulness of the treatment and collaboration, whereas therapists are more concerned with the performance of the client and their own confidence as therapists [64].

Patients' interim and posttreatment HAQ scores were positively associated with treatment outcomes. The final interim model with the HAQ-Helpfulness scores of patients as the sole predictor explained 9.8% (F_1=17.03; *P*<.001) of the variance in eating disorder pathology. Post treatment, the HAQ-Helpfulness score of patients explained 22.3% (F_1=43.58; *P*<.001) of the variance in eating disorder pathology. This corresponds with the results of other studies [27,30,65]. Although the explained variance is not that high, meaning that there are more unknown factors influencing eating disorder pathology, it does show that patients who are confident in their own capacity to improve their situation are more likely to have better treatment outcomes. This is in line with a narrative review of web-based interventions that showed that in most studies, helpfulness-related factors were found to be positively associated with treatment outcomes in internet interventions [33,66]. In only one of the studies described in the narrative review, a positive association between cooperation-related factors and treatment outcomes was found [33]. Patients who opt for web-CBT may value cooperation with a therapist less important

than patients who prefer face-to-face treatment and may prefer the relative anonymity of the internet [64]. For web-CBT for ED, we found one study that also reported a positive association between TA and treatment outcomes [65]. However, this study used a different measure to operationalize the construct of TA and did not focus on the perspective of therapists.

Finally, treatment completion is an important predictor of treatment outcome [41,66,67]. In a previous study by our laboratory [41], we found that completers had significantly better treatment outcomes than noncompleters. This highlights the importance of investigating predictors of web-CBT treatment completion.

Incidentally, we found that patients with higher BMI completed treatment more often. This is in contrast to observations by Werz et al [28]. On the basis of the current scientific knowledge, we have no reason to interpret our findings as clinically relevant.

This study reported a positive association between TA ratings and treatment completion. More specifically, the univariate models indicated that all HAQ scales (helpfulness, cooperation, and the total score), as scored by both patients and therapists, were predictors of treatment completion both at interim and post treatment. However, the multivariate model indicated that only the patients' HAQ-total score and therapists' HAQ-Helpfulness score were positively associated with treatment completion. This might indicate that treatment noncompletion could be reduced by improving the TA.

It should be noted that the results of this study are limited by a lack of consensus within the field of TA research concerning the definition and operationalization of TA. Across studies, a wide diversity of measures, such as the Working Alliance Inventory [67] and Therapeutic Alliance Scale [68], are used to operationalize the TA and were designed for face-to-face treatments [37]. This reduces the cross-comparability between studies. Establishing a consensus concerning the operationalization of the construct of TA in EDs and other psychotherapeutic treatments, specifically focusing on web-based treatment, would therefore be very welcome. It should also be noted that the HAQ does not provide norm scores regarding the quality of the TA, which makes it difficult to determine whether the TA is good. No clinically relevant differences in TA were determined.

Owing to the rapid development of mobile- and internet-based technology, tools and apps that are integrated into mobile devices such as smartphones are increasingly being used. The data of this study were collected from 2011 to 2013 and have already shown the importance of investing in TA because it could contribute to less psychopathology and more treatment completion. With the increased options for interactivity, it is becoming increasingly interesting to study the impact of TA in web-based treatment.

For future studies, we suggest including a more extensive population because it could lead to different results and insights. For example, male patients with ED are increasingly recognized and have unique concerns regarding disordered eating and body image [69]. The same applies to patients with anorexia nervosa. In this population, high dropout rates have been reported [70], and the strength of the TA has been shown to be associated with changes in ED symptoms [71,72].

It would also be interesting to include a face-to-face CBT condition, as this allows a comparative estimation of the effectiveness of TA on treatment outcome and treatment completion.

Finally, monitoring TA from the patient's perspective and acting on relatively low and diminishing scores throughout the treatment process might be fruitful for clinical practice and contribute to better treatment results and completion.

Conclusions

The results of this study showed that the strength of the TA during web-CBT for ED increased for patients who completed the program and decreased for patients who did not from both the perspectives of patients and therapists. Our study also showed that TA is predictive of ED pathology and treatment completion. In particular, patients' confidence in their own abilities, measured using the HAQ-Helpfulness subscale, is important for predicting posttreatment ED pathology and treatment completion.

Conflicts of Interest

None declared.

References

1. Agras WS. The consequences and costs of the eating disorders. Psychiatr Clin North Am 2001 Jun;24(2):371-379. [doi: 10.1016/s0193-953x(05)70232-x] [Medline: 11416936]
2. Preti A, Girolamo GD, Vilagut G, Alonso J, Graaf RD, Bruffaerts R, ESEMeD-WMH Investigators. The epidemiology of eating disorders in six European countries: results of the ESEMeD-WMH project. J Psychiatr Res 2009 Sep;43(14):1125-1132. [doi: 10.1016/j.jpsychires.2009.04.003] [Medline: 19427647]
3. Ali K, Farrer L, Fassnacht DB, Gulliver A, Bauer S, Griffiths KM. Perceived barriers and facilitators towards help-seeking for eating disorders: a systematic review. Int J Eat Disord 2017 Jan;50(1):9-21. [doi: 10.1002/eat.22598] [Medline: 27526643]
4. Hart LM, Granillo MT, Jorm AF, Paxton SJ. Unmet need for treatment in the eating disorders: a systematic review of eating disorder specific treatment seeking among community cases. Clin Psychol Rev 2011 Jul;31(5):727-735. [doi: 10.1016/j.cpr.2011.03.004] [Medline: 21501580]

5. J Banasiak S, J Paxton S, J Hay P. Evaluating accessible treatments for bulimic eating disorders in primary care. Aust J Prim Health 1998;4(3):147. [doi: 10.1071/PY98042]
6. Becker AE, Hadley Arrindell A, Perloe A, Fay K, Striegel-Moore RH. A qualitative study of perceived social barriers to care for eating disorders: perspectives from ethnically diverse health care consumers. Int J Eat Disord 2010 Nov 01;43(7):633-647 [FREE Full text] [doi: 10.1002/eat.20755] [Medline: 19806607]
7. Cachelin FM, Striegel-Moore RH. Help seeking and barriers to treatment in a community sample of Mexican American and European American women with eating disorders. Int J Eat Disord 2006 Mar;39(2):154-161. [doi: 10.1002/eat.20213] [Medline: 16252278]
8. Evans EJ, Hay PJ, Mond J, Paxton SJ, Quirk F, Rodgers B, et al. Barriers to help-seeking in young women with eating disorders: a qualitative exploration in a longitudinal community survey. Eat Disord 2011 Apr 22;19(3):270-285. [doi: 10.1080/10640266.2011.566152] [Medline: 21516551]
9. Schmidt U. Getting technical. Eur Eat Disorders Rev 2003 May;11(3):147-154. [doi: 10.1002/erv.522]
10. Agras WS, Fitzsimmons-Craft EE, Wilfley DE. Evolution of cognitive-behavioral therapy for eating disorders. Behav Res Ther 2017 Jan;88:26-36 [FREE Full text] [doi: 10.1016/j.brat.2016.09.004] [Medline: 28110674]
11. Barakat S, Maguire S, Smith KE, Mason TB, Crosby RD, Touyz S. Evaluating the role of digital intervention design in treatment outcomes and adherence to eTherapy programs for eating disorders: a systematic review and meta-analysis. Int J Eat Disord 2019 Oct;52(10):1077-1094. [doi: 10.1002/eat.23131] [Medline: 31328815]
12. Fitzsimmons-Craft EE, Taylor CB, Graham AK, Sadeh-Sharvit S, Balantekin KN, Eichen DM, et al. Effectiveness of a digital cognitive behavior therapy-guided self-help intervention for eating disorders in college women: a cluster randomized clinical trial. JAMA Netw Open 2020 Aug 03;3(8):e2015633 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.15633] [Medline: 32865576]
13. Jenkins PE, Luck A, Violato M, Robinson C, Fairburn CG. Clinical and cost-effectiveness of two ways of delivering guided self-help for people with an eating disorder: a multi-arm randomized controlled trial. Int J Eat Disord 2021 Jul;54(7):1224-1237. [doi: 10.1002/eat.23554] [Medline: 33998020]
14. Jensen E, Linnet J, Holmberg T, Tarp K, Nielsen J, Lichtenstein M. Effectiveness of internet-based guided self-help for binge-eating disorder and characteristics of completers versus noncompleters. Int J Eat Disord 2020 Dec;53(12):2026-2031 [FREE Full text] [doi: 10.1002/eat.23384] [Medline: 32918321]
15. Machado PP, Rodrigues TF. Treatment delivery strategies for eating disorders. Curr Opin Psychiatry 2019 Nov;32(6):498-503. [doi: 10.1097/YCO.0000000000000542] [Medline: 31306248]
16. Moghimi E, Davis C, Rotondi M. The efficacy of eHealth interventions for the treatment of adults diagnosed with full or subthreshold binge eating disorder: systematic review and meta-analysis. J Med Internet Res 2021 Jul 20;23(7):e17874 [FREE Full text] [doi: 10.2196/17874] [Medline: 34283028]
17. Schuster R, Berger T, Laireiter A. Computer und Psychotherapie – geht das zusammen? Psychotherapeut 2017 Aug 3;63(4):271-282. [doi: 10.1007/s00278-017-0214-8]
18. Taylor C, Graham A, Flatt R, Waldherr K, Fitzsimmons-Craft E. Current state of scientific evidence on internet-based interventions for the treatment of depression, anxiety, eating disorders and substance abuse: an overview of systematic reviews and meta-analyses. Eur J Public Health 2021 Jul 07;31(31 Suppl 1):i3-10 [FREE Full text] [doi: 10.1093/eurpub/ckz208] [Medline: 32918448]
19. ter Huurne ED, de Haan HA, Postel MG, van der Palen J, VanDerNagel JE, DeJong CA. Web-based cognitive behavioral therapy for female patients with eating disorders: randomized controlled trial. J Med Internet Res 2015 Jun 18;17(6):e152 [FREE Full text] [doi: 10.2196/jmir.3946] [Medline: 26088580]
20. Wyssen A, Meyer AH, Messerli-Bürgy N, Forrer F, Vanhulst P, Lalanne D, et al. BED-online: acceptance and efficacy of an internet-based treatment for binge-eating disorder: a randomized clinical trial including waitlist conditions. Eur Eat Disord Rev 2021 Nov;29(6):937-954. [doi: 10.1002/erv.2856] [Medline: 34418221]
21. Vollert B, Beintner I, Musiat P, Gordon G, Görlich D, Nacke B, et al. Using internet-based self-help to bridge waiting time for face-to-face outpatient treatment for Bulimia Nervosa, Binge Eating Disorder and related disorders: study protocol of a randomized controlled trial. Internet Interv 2019 Apr;16:26-34 [FREE Full text] [doi: 10.1016/j.invent.2018.02.010] [Medline: 30775262]
22. Yim SH, Schmidt U. Experiences of computer-based and conventional self-help interventions for eating disorders: a systematic review and meta-synthesis of qualitative research. Int J Eat Disord 2019 Oct;52(10):1108-1124. [doi: 10.1002/eat.23142] [Medline: 31343088]
23. Nienhuis JB, Owen J, Valentine JC, Winkeljohn Black S, Halford TC, Parazak SE, et al. Therapeutic alliance, empathy, and genuineness in individual adult psychotherapy: a meta-analytic review. Psychother Res 2018 Jul;28(4):593-605. [doi: 10.1080/10503307.2016.1204023] [Medline: 27389666]
24. Horvath AO, Del Re AC, Flückiger C, Symonds D. Alliance in individual psychotherapy. Psychotherapy (Chic) 2011 Mar;48(1):9-16. [doi: 10.1037/a0022186] [Medline: 21401269]
25. Flückiger C, Del Re AC, Wampold BE, Horvath AO. The alliance in adult psychotherapy: a meta-analytic synthesis. Psychotherapy (Chic) 2018 Dec;55(4):316-340. [doi: 10.1037/pst0000172] [Medline: 29792475]

26. Norcross JC, Lambert MJ. Psychotherapy relationships that work III. Psychotherapy (Chic) 2018 Dec;55(4):303-315. [doi: 10.1037/pst0000193] [Medline: 30335448]
27. Probst G, Berger T, Flückiger C. Die allianz als prädiktor für den therapieerfolg internetbasierter interventionen bei psychischen störungen: eine korrelative metaanalyse. Verhaltenstherapie 2019 Jul 17;29(3):182-195. [doi: 10.1159/000501565]
28. Werz J, Voderholzer U, Tuschen-Caffier B. Alliance matters: but how much? A systematic review on therapeutic alliance and outcome in patients with anorexia nervosa and bulimia nervosa. Eat Weight Disord 2022 May;27(4):1279-1295 [FREE Full text] [doi: 10.1007/s40519-021-01281-7] [Medline: 34374966]
29. Puls H, Schmidt R, Hilbert A. Therapist adherence and therapeutic alliance in individual cognitive-behavioural therapy for adolescent binge-eating disorder. Eur Eat Disord Rev 2019 Mar;27(2):182-194. [doi: 10.1002/erv.2650] [Medline: 30334340]
30. Graves TA, Tabri N, Thompson-Brenner H, Franko DL, Eddy KT, Bourion-Bedes S, et al. A meta-analysis of the relation between therapeutic alliance and treatment outcome in eating disorders. Int J Eat Disord 2017 Apr 02;50(4):323-340. [doi: 10.1002/eat.22672] [Medline: 28152196]
31. Andersson G, Paxling B, Wiwe M, Vernmark K, Felix CB, Lundborg L, et al. Therapeutic alliance in guided internet-delivered cognitive behavioural treatment of depression, generalized anxiety disorder and social anxiety disorder. Behav Res Ther 2012 Sep;50(9):544-550. [doi: 10.1016/j.brat.2012.05.003] [Medline: 22728647]
32. Andersson G, Cuijpers P, Carlbring P, Riper H, Hedman E. Guided internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: a systematic review and meta-analysis. World Psychiatry 2014 Oct;13(3):288-295 [FREE Full text] [doi: 10.1002/wps.20151] [Medline: 25273302]
33. Berger T. The therapeutic alliance in internet interventions: a narrative review and suggestions for future research. Psychother Res 2017 Sep;27(5):511-524. [doi: 10.1080/10503307.2015.1119908] [Medline: 26732852]
34. Clarke J, Proudfoot J, Whitton A, Birch M, Boyd M, Parker G, et al. Therapeutic alliance with a fully automated mobile phone and web-based intervention: secondary analysis of a randomized controlled trial. JMIR Ment Health 2016 Feb 25;3(1):e10 [FREE Full text] [doi: 10.2196/mental.4656] [Medline: 26917096]
35. Simpson SG, Reid CL. Therapeutic alliance in videoconferencing psychotherapy: a review. Aust J Rural Health 2014 Dec;22(6):280-299. [doi: 10.1111/ajr.12149] [Medline: 25495622]
36. Steel K, Cox D, Garry H. Therapeutic videoconferencing interventions for the treatment of long-term conditions. J Telemed Telecare 2011;17(3):109-117. [doi: 10.1258/jtt.2010.100318] [Medline: 21339304]
37. Sucala M, Schnur JB, Constantino MJ, Miller SJ, Brackman EH, Montgomery GH. The therapeutic relationship in e-therapy for mental health: a systematic review. J Med Internet Res 2012 Aug 02;14(4):e110 [FREE Full text] [doi: 10.2196/jmir.2084] [Medline: 22858538]
38. Schlegl S, Bürger C, Schmidt L, Herbst N, Voderholzer U. The potential of technology-based psychological interventions for anorexia and bulimia nervosa: a systematic review and recommendations for future research. J Med Internet Res 2015 Mar 31;17(3):e85 [FREE Full text] [doi: 10.2196/jmir.3554] [Medline: 25840591]
39. Ertelt TW, Crosby RD, Marino JM, Mitchell JE, Lancaster K, Crow SJ. Therapeutic factors affecting the cognitive behavioral treatment of bulimia nervosa via telemedicine versus face-to-face delivery. Int J Eat Disord 2011 Dec;44(8):687-691 [FREE Full text] [doi: 10.1002/eat.20874] [Medline: 22072405]
40. Hamatani S, Numata N, Matsumoto K, Sutoh C, Ibuki H, Oshiro K, et al. Internet-based cognitive behavioral therapy via videoconference for patients with bulimia nervosa and binge-eating disorder: pilot prospective single-arm feasibility trial. JMIR Form Res 2019 Oct 23;3(4):e15738 [FREE Full text] [doi: 10.2196/15738] [Medline: 31647472]
41. Ter Huurne ED, Postel MG, de Haan HA, van der Palen J, DeJong CA. Treatment dropout in web-based cognitive behavioral therapy for patients with eating disorders. Psychiatry Res 2017 Jan;247:182-193. [doi: 10.1016/j.psychres.2016.11.034] [Medline: 27918968]
42. Laws HB, Constantino MJ, Sayer AG, Klein DN, Kocsis JH, Manber R, et al. Convergence in patient-therapist therapeutic alliance ratings and its relation to outcome in chronic depression treatment. Psychother Res 2017 Jul;27(4):410-424 [FREE Full text] [doi: 10.1080/10503307.2015.1114687] [Medline: 26829714]
43. Tschuschke V, Koemeda-Lutz M, von Wyl A, Crameri A, Schulthess P. The impact of patients' and therapists' views of the therapeutic alliance on treatment outcome in psychotherapy. J Nerv Ment Dis 2020 Jan;208(1):56-64. [doi: 10.1097/NMD.0000000000001111] [Medline: 31688492]
44. ter Huurne ED, Postel MG, de Haan HA, DeJong CA. Effectiveness of a web-based treatment program using intensive therapeutic support for female patients with bulimia nervosa, binge eating disorder and eating disorders not otherwise specified: study protocol of a randomized controlled trial. BMC Psychiatry 2013 Nov 16;13:310 [FREE Full text] [doi: 10.1186/1471-244X-13-310] [Medline: 24238630]
45. Ter Huurne ED, de Haan HA, Postel MG, DeJong CAJ, VanDerNagel JEL, van der Palen J. Long-term effectiveness of web-based cognitive behavioral therapy for patients with eating disorders. Eat Weight Disord 2021 Apr;26(3):911-919. [doi: 10.1007/s40519-020-00929-0] [Medline: 32449152]
46. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59 Suppl 20:22-33;quiz 34. [Medline: 9881538]

47. Overbeek I, Schruers K, Griez E, Overbeek JM. Mini International Neuropsychiatric Interview: Nederlandse Versie 5.0.0, DSM-IV [Dutch Version]. Mini International Neuropsychiatric Interview: Nederlandse Versie 5.0.0, DSM-IV [Dutch Version] 1999 [FREE Full text]
48. Brownley KA, Berkman ND, Sedway JA, Lohr KN, Bulik CM. Binge eating disorder treatment: a systematic review of randomized controlled trials. Int J Eat Disord 2007 May;40(4):337-348. [doi: 10.1002/eat.20370] [Medline: 17370289]
49. Shapiro JR, Berkman ND, Brownley KA, Sedway JA, Lohr KN, Bulik CM. Bulimia nervosa treatment: a systematic review of randomized controlled trials. Int J Eat Disord 2007 May;40(4):321-336. [doi: 10.1002/eat.20372] [Medline: 17370288]
50. Wilson GT, Grilo CM, Vitousek KM. Psychological treatment of eating disorders. Am Psychol 2007 Apr;62(3):199-216. [doi: 10.1037/0003-066X.62.3.199] [Medline: 17469898]
51. Britt E, Hudson SM, Blampied NM. Motivational interviewing in health settings: a review. Patient Educ Couns 2004 May;53(2):147-155. [doi: 10.1016/S0738-3991(03)00141-1] [Medline: 15140454]
52. Miller JH, Moyers T. Motivational interviewing in substance abuse: applications for occupational medicine. Occup Med 2002;17(1):51-65, iv. [Medline: 11726336]
53. Luborsky L, McLellan AT, Woody GE, O'Brien CP, Auerbach A. Therapist success and its determinants. Arch Gen Psychiatry 1985 Jun;42(6):602-611. [doi: 10.1001/archpsyc.1985.01790290084010] [Medline: 4004503]
54. De Weert-Van Oene GH, De Jong CA, Jörg F, Schrijvers GJ. The Helping Alliance Questionnaire: psychometric properties in patients with substance dependence. Subst Use Misuse 1999 Sep;34(11):1549-1569. [doi: 10.3109/10826089909039414] [Medline: 10468107]
55. The Penn helping alliance scales. In: The Psychotherapeutic Process A Research Handbook. New York: Guilford Publications; 1986.
56. Fairburn C, Beglin S. Assessment of eating disorders: interview or self-report questionnaire? Int J Eat Disord 1994 Dec;16(4):363-370. [Medline: 7866415]
57. Byrne SM, Allen KL, Lampard AM, Dove ER, Fursland A. The factor structure of the eating disorder examination in clinical and community samples. Int J Eat Disord 2010 Apr;43(3):260-265. [doi: 10.1002/eat.20681] [Medline: 19350647]
58. Black CM, Wilson GT. Assessment of eating disorders: interview versus questionnaire. Int J Eat Disord 1996 Jul;20(1):43-50. [doi: 10.1002/(SICI)1098-108X(199607)20:1<43::AID-EAT5>3.0.CO;2-4] [Medline: 8807351]
59. Eating disorder examination Questionnaire (Edition 6.0D). In: Cognitive Behavior Therapy and Eating Disorders. New York: Guilford Press; 2008.
60. IBM SPSS Statistics for Windows, Version 20.0. Armonk, NY: IBM Corp; 2013.
61. Cohen J. The effect size. In: Statistical Power Analysis for the Behavioral Sciences. New York: Routledge; 1988.
62. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016 Jun;15(2):155-163 [FREE Full text] [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]
63. Tschuschke V, Crameri A, Koehler M, Berglar J, Muth K, Staczan P, et al. The role of therapists' treatment adherence, professional experience, therapeutic alliance, and clients' severity of psychological problems: prediction of treatment outcome in eight different psychotherapy approaches. Preliminary results of a naturalistic study. Psychother Res 2015 Apr 15;25(4):420-434. [doi: 10.1080/10503307.2014.896055] [Medline: 24689912]
64. Leibert T, Archer J, Munson J, York G. An exploratory study of client perceptions of internet counseling and the therapeutic alliance. J Mental Health Counsel 2006;28(1):69-83 [FREE Full text] [doi: 10.17744/mehc.28.1.f0h37djrw89nv6vb]
65. Dölemeyer R, Klinitzke G, Steinig J, Wagner B, Kersting A. [Working alliance in internet-based therapy for binge eating disorder]. Psychiatr Prax 2013 Sep;40(6):321-326. [doi: 10.1055/s-0033-1343320] [Medline: 24008680]
66. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152. [doi: 10.2196/jmir.2104] [Medline: 23151820]
67. Horvath A, Greenberg L. The development of the working alliance inventory. In: The Psychotherapeutic Process: A Research Handbook. New York: Guilford Press; 1986.
68. Bickman L, Vides de Andrade AR, Warren Lambert E, Doucette A, Sapyta J, Suzanne Boyd A, et al. Youth therapeutic alliance in intensive treatment settings. J Behav Health Services Res 2004 Apr;31(2):134-148. [doi: 10.1007/bf02287377]
69. Nagata JM, Ganson KT, Murray SB. Eating disorders in adolescent boys and young men: an update. Curr Opin Pediatr 2020 Aug;32(4):476-481 [FREE Full text] [doi: 10.1097/MOP.0000000000000911] [Medline: 32520822]
70. Lock J, Brandt H, Woodside B, Agras S, Halmi WK, Johnson C, et al. Challenges in conducting a multi-site randomized clinical trial comparing treatments for adolescent anorexia nervosa. Int J Eat Disord 2012 Mar;45(2):202-213 [FREE Full text] [doi: 10.1002/eat.20923] [Medline: 21495052]
71. Isserlin L, Couturier J. Therapeutic alliance and family-based treatment for adolescents with anorexia nervosa. Psychotherapy (Chic) 2012 Mar;49(1):46-51. [doi: 10.1037/a0023905] [Medline: 21967072]
72. Pereira T, Lock J, Oggins J. Role of therapeutic alliance in family therapy for adolescent anorexia nervosa. Int J Eat Disord 2006 Dec;39(8):677-684. [doi: 10.1002/eat.20303] [Medline: 16937386]

Abbreviations

BED: binge eating disorder
BN: bulimia nervosa
CBT: cognitive behavioral therapy
DSM-IV: Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition
ED: eating disorder
EDE-Q: Eating Disorder Examination Questionnaire
EDNOS: eating disorder not otherwise specified
HAQ: Helping Alliance Questionnaire
ICC: intraclass correlation coefficient
OR: odds ratio
TA: therapeutic alliance

Edited by A Mavragani; submitted 24.09.21; peer-reviewed by S Chokshi, B Taylor; comments to author 31.12.21; revised version received 20.04.22; accepted 24.04.22; published 30.06.22.

Please cite as:
Stoeten C, de Haan HA, Postel MG, Brusse-Keizer M, ter Huurne ED
Therapeutic Alliance in Web-Based Treatment for Eating Disorders: Secondary Analysis of a Randomized Controlled Trial
JMIR Form Res 2022;6(6):e33813
URL: https://formative.jmir.org/2022/6/e33813
doi:10.2196/33813
PMID:35771608

Original Paper

Development of a WeChat-based Mobile Messaging Smoking Cessation Intervention for Chinese Immigrant Smokers: Qualitative Interview Study

Nan Jiang[1], PhD; Erin S Rogers[1], MPH, DrPH; Paula Cupertino[2], PhD; Xiaoquan Zhao[3], PhD; Francisco Cartujano-Barrera[2], MD; Joanne Chen Lyu[4], PhD; Lu Hu[1], PhD; Scott E Sherman[1,5], MD, MPH

[1]Department of Population Health, Grossman School of Medicine, New York University, New York, NY, United States

[2]Wilmot Cancer Institute, University of Rochester Medical Center, Rochester, NY, United States

[3]Department of Communication, College of Humanities and Social Sciences, George Mason University, Fairfax, VA, United States

[4]Center for Tobacco Control Research and Education, University of California, San Francisco, CA, United States

[5]Department of Medicine, Veterans Affairs New York Harbor Healthcare System, New York, NY, United States

Corresponding Author:
Nan Jiang, PhD
Department of Population Health
Grossman School of Medicine
New York University
180 Madison Ave, Rm #17-54
New York, NY, 10016
United States
Phone: 1 646 501 3553
Email: Nan.Jiang@nyulangone.org

Abstract

Background: Smoking remains a major public health issue among Chinese immigrants. Smoking cessation programs that focus on this population are scarce and have a limited population-level impact due to their low reach. Mobile messaging interventions have the potential to reach large audiences and expand smokers' access to smoking cessation treatment.

Objective: This study describes the development of a culturally and linguistically appropriate mobile messaging smoking cessation intervention for Chinese immigrant smokers delivered via WeChat, the most frequently used social media platform among Chinese people globally.

Methods: This study had 2 phases. In phase 1, we developed a mobile message library based on social cognitive theory and the US Clinical Practice Guidelines for Treating Tobacco Use and Dependence. We culturally adapted messages from 2 social cognitive theory-based text messaging smoking cessation programs (SmokefreeTXT and Decídetexto). We also developed new messages targeting smokers who were not ready to quit smoking and novel content addressing Chinese immigrant smokers' barriers to quitting and common misconceptions related to willpower and nicotine replacement therapy. In phase 2, we conducted in-depth interviews with 20 Chinese immigrant smokers (including 7 women) in New York City between July and August 2021. The interviews explored the participants' smoking and quitting experiences followed by assessment of the text messages. Participants reviewed 17 text messages (6 educational messages, 3 self-efficacy messages, and 8 skill messages) via WeChat and rated to what extent the messages enhanced their motivation to quit, promoted confidence in quitting, and increased awareness about quitting strategies. The interviews sought feedback on poorly rated messages, explored participant preferences for content, length, and format, discussed their concerns with WeChat cessation intervention, and solicited recommendations for frequency and timing of messages.

Results: Overall, participants reported that the messages enhanced their motivation to quit, offered encouragement, and made them more informed about how to quit. Participants particularly liked the messages about the harms of smoking and strategies for quitting. They reported barriers to applying some of the quitting strategies, including coping with stress and staying abstinent at work. Participants expressed strong interest in the WeChat mobile messaging cessation intervention and commented on its potential to expand their access to smoking cessation treatment.

Conclusions: Mobile messages are well accepted by Chinese immigrant smokers. Research is needed to assess the feasibility, acceptability, and efficacy of WeChat mobile messaging smoking cessation interventions for promoting abstinence among Chinese immigrant smokers.

(JMIR Form Res 2022;6(6):e36091) doi:10.2196/36091

KEYWORDS

smoking cessation; text messaging; mobile health; Chinese American

Introduction

Despite the considerable decline in smoking prevalence in the United States over the past 50 years [1], smoking remains a major public health issue among Chinese American immigrants, particularly men. In New York City, for example, the current (past 30-day) smoking rate among the general male population is 17.5%, whereas the rate for Chinese American men is 28.2% [2]. Foreign-born Chinese Americans are more likely to smoke than those who are born in the United States [2]. The current smoking rate among Chinese American women remains low (1.9%) compared to the general New York City female population (9.2%) [2], reflecting the traditional Chinese cultural norm against women smoking [3].

Culturally and linguistically appropriate smoking cessation programs for Chinese Americans are scarce [4-6]. Moreover, cessation programs targeting this population with demonstrated efficacy, such as the Asian Smokers' Quitline [7] and community-based cessation programs [8-10], often have limited reach. Utilization of the Asian Smokers' Quitline is low (1.3%), with an estimated 2010 Chinese-, Korean-, and Vietnamese-speaking smokers using the Asian Smokers' Quitline every year [11]. This low reach has limited the population impact of the cessation program.

Chinese immigrant smokers face a range of barriers to using smoking cessation programs. These include low awareness of existing programs, skepticism about treatment effectiveness, and time constraints [12-17]. All these factors have deterred Chinese immigrant smokers from using the quitline or local cessation programs. Thus, addressing smoking among Chinese immigrants requires scalable interventions that are engaging, accessible, and efficacious to improve abstinence rates. In addition, Chinese immigrant smokers are largely not ready to quit, with only 6% to 33% planning to quit within a month [18-20]. This is attributed to multiple factors, including low awareness of the harms of smoking and attachment to traditional Chinese norms that support men smoking [12-14,20-23]. Hence, smoking cessation programs targeting Chinese immigrant populations must engage a broad group of smokers, including those who are not ready to quit (ie, those who have no quit date in the next 30 days), in order to optimize the population impact.

Short message service (SMS) cessation programs, which can reach large audiences and expand smokers' access to treatment, have been shown to be effective in increasing long-term abstinence [24-26]. Current SMS cessation programs, including the National Cancer Institute's SmokefreeTXT [27] and most clinical trials [26], primarily focus on smokers who are ready to quit (ie, smokers who have a quit date within a month) and exclude unmotivated smokers, who represent a significant portion of smokers. Little is known about the feasibility and treatment efficacy of SMS programs in engaging unmotivated smokers.

To fill this research gap, we are conducting a study to examine the feasibility, acceptability, and preliminary effectiveness of a culturally adapted, linguistically appropriate WeChat-based mobile messaging smoking cessation intervention among Chinese immigrant smokers. WeChat is a social networking application widely used among Chinese people globally [28]. In a pilot study with Chinese immigrant smokers in New York City, we found that WeChat was more frequently used than SMS or other social media sites, such as Facebook, WhatsApp, and Twitter [12]. As part of our larger study, we have developed a theory-based mobile message library tailored to Chinese immigrant smokers. We conducted in-depth interviews to assess content preferences and solicit suggestions to guide further message modifications. In this paper, we present the process of message development and adaptation, including findings from the interviews.

Methods

Following a guide for developing SMS programs for health behaviors [29], this study consisted of two phases: (1) message development and adaptation and (2) in-depth interviews.

Phase 1: Development and Adaptation of Mobile Messages

We developed a mobile message library based on social cognitive theory (SCT) and the US Clinical Practice Guidelines for Treating Tobacco Use and Dependence [30]. SCT has been widely used in SMS smoking cessation interventions as a conceptual framework [31-36]. Our messages were designed around key SCT constructs to promote the motivation to quit, increase knowledge and skills related to quitting, improve quitting self-efficacy, and address social norms and misconceptions related to smoking and quitting. Specifically, the content described the adverse health effects of smoking, including light and social smoking, and the harms of exposure to secondhand and thirdhand smoke (ie, behavioral capability); introduced reasons for and benefits of quitting (ie, outcome expectations); encouraged smokers to make quit attempts (ie, self-efficacy); offered advice on quitting preparation (eg, redesigning the environment, knowing smoking triggers, setting a quit date, and securing social support from family and friends) and problem troubleshooting (ie, behavioral capability); provided cognitive and behavioral strategies, such as refusal and coping skills; addressed myths about willpower; described the effect of nicotine replacement therapy; highlighted the

prevalence of smoking among Chinese Americans (ie, social norms); and provided information about smoking cessation programs targeting Chinese American smokers.

Our library included messages drawn from 2 SCT-based SMS smoking cessation programs focusing on smokers who are ready to quit, including SmokefreeTXT and Decídetexto (a program designed for Latino smokers [37,38]). We adapted the messages to the cultural context of Chinese immigrant smokers. For example, coping strategies were adapted to include practices relevant to Chinese immigrants (eg, instead of "watching a movie and enjoying a handful of popcorn," as suggested by the SmokefreeTXT, we suggested the following: "tidy your home/workplace for a few minutes" and "call or text your spouse or friends").

In addition, we developed new messages targeting smokers not ready to quit in the next 30 days. The content focused on addressing Chinese immigrant smokers' misconceptions about smoking and quitting (eg, misconceptions related to willpower and nicotine replacement therapy) and common challenges in quitting [12-17,39,40]. New messages were created based on SCT, along with the recommendations of the US Clinical Practice Guidelines for smoking cessation treatment for unmotivated smokers. The new messages included (1) more details about the health hazards of smoking, including light and social smoking (ie, behavioral capability), (2) more motivational and encouraging content (ie, self-efficacy), including passages such as the following: "In your journey of quitting, there may be little stumbles. It's the progress you are making to achieve success. As long as you get up, it's not failure. Most former smokers have tried several times before they finally quit. Each quit attempt, even it doesn't work, you may learn something new about yourself and therefore, you're a step closer to becoming a former smoker," (3) more concrete advice on quitting and problem troubleshooting (ie, behavioral capability), (4) emphasis on the smoke-free lifestyle valued by most Chinese Americans (ie, social norms), including information such as the following: "The majority of Chinese immigrants living in New York City do not smoke. This includes 72% of Chinese American men and 98% of Chinese American women," (5) encouragement to use the Asian Smokers' Quitline and local language-specific cessation programs (ie, behavioral capability), and (6) myths about willpower and nicotine replacement therapy (ie, behavioral capability).

Messages were developed and adapted through an iterative process. The first author (NJ), a bilingual researcher with English and Chinese proficiency, drafted initial messages in English. Other authors (SES and ESR) reviewed and commented on the messages. Once finalized, all messages were translated into Chinese by the first author and reviewed by another bilingual author (XZ). The messages were not tailored to gender, because for Chinese immigrant smokers, men and women experience similar barriers to cessation (eg, lack of knowledge and quitting skills) and share similar misconceptions (eg, related to willpower) [12].

Phase 2. In-depth Interviews

Participants and Recruitment

Between July and August, 2021, we recruited 20 Chinese immigrant smokers (including men and women) by posting flyers in a community-based organization that primarily serves Chinese Americans and via in-person contacts (ie, snowball sampling). The eligibility criteria included the following: (1) age 18 to 65 years, (2) self-identification as a Chinese immigrant, (3) smoking history of at least 100 total lifetime cigarettes, (4) currently smoking at least 3 days per week, (5) current use of WeChat at least 3 days per week, (6) ownership of a smartphone, (7) ability to speak and read Chinese, (8) residence in New York City, and (9) being somewhat interested in quitting smoking, as assessed by the following screening question: "Which statement best describes your intention to quit smoking?" with response options including "I don't want to quit at all," "I may quit at some point, but not in the next 6 months," "I plan to quit within the next 6 months," "I plan to quit within the next 30 days," and "I am currently trying to quit." Subjects were excluded if they answered "I don't want to quit at all." Other exclusion criteria included participation in other smoking cessation services and being pregnant or breastfeeding.

In-depth Interview Procedures

Semistructured in-depth interviews were conducted in person. Two research staff moderated each interview in Chinese. Prior to the interview, participants signed written consent, connected with one research staff through WeChat, and completed a brief paper-and-pencil survey of their demographic information, smoking patterns, and WeChat use frequency.

The interviews started with questions about participants' smoking and quitting experiences and plans for quitting, followed by assessment of the messages. A total of 17 text messages were tested, including 6 educational messages, 3 self-efficacy messages, and 8 skill messages (Table 1). We selected messages that represented various SCT constructs and were relatively long, compared with other messages in the library. The messages were text only, with no videos or pictures. Our research staff sent the 17 text messages to the participants via WeChat during the interview. Participants read each message on their phone and rated them on a rating sheet with a 0 to 10 visual analog scale to indicate to what extent the message enhanced their motivation to quit (for the educational messages), promoted their confidence in quitting (for the self-efficacy messages), and increased their awareness of quitting strategies (for the skill messages). Higher scores indicated higher levels of motivation, confidence, and awareness. The ratings were solely used to facilitate the discussions (eg, to gather feedback on the poorly rated messages), rather than to make comparisons across participants or messages. After the rating was completed, we asked participants what they liked and did not like about the messages, and whether the contents were relevant to them. We gathered feedback on the poorly rated messages to gain insights about content preferences and sought feedback on message length and format. We explored the meanings of key concepts (eg, how participants described substitute behaviors), explored examples of coping and refusal strategies, discussed concerns

about using WeChat for smoking cessation intervention, and elicited suggestions for messaging frequency and timing. We also asked participants to summarize each message in order to assess their understanding of the content. The interviews typically lasted between 60 and 90 minutes. The interviews were audio-recorded and transcribed into Chinese. Participants were compensated with a US $75 gift card.

Table 1. Text messages assessed in the interview.

Content of messages	Social cognitive theory construct	Number of messages (total=17)
Educational messages		
Harms of smoking	Behavioral capability	2
Health benefits of quitting	Outcome expectations	1
Reasons for quitting	Outcome expectations	1
Key elements in quitting	Behavioral capability	1
Social and light smoking	Behavioral capability	1
Self-efficacy messages	Self-efficacy	3
Quitting skill messages		
Redesign environment	Behavioral capability	1
Identify smoking triggers	Behavioral capability	1
Explore substitute behaviors to cope with cravings	Behavioral capability	2
Refusal strategies	Behavioral capability	1
Secure social support from family and friends	Behavioral capability	1
Set a quit date	Behavioral capability	1
Misconception: willpower	Behavioral capability	1

Data Analysis

The data were analyzed with NVivo version 12 (QSR International). The first author closely read all transcripts (in Chinese) and created an initial codebook of themes and subthemes in 3 domains related to the research questions (eg, quitting experience, perceptions about the messages, and attitudes toward the WeChat mobile messaging smoking cessation intervention). Two coders (NJ and another bilingual team member) independently coded a subset of transcripts, generated emergent themes and subthemes using an inductive analytic approach [41], discussed and resolved disagreements on the codes, and finalized the codebook. The first author then coded the remaining transcripts and selected illustrative quotes. The quotes were reviewed by both coders and translated into English.

Ethics Approval

The study protocol was approved by the Institutional Review Board of New York University Grossman School of Medicine (i20-01959).

Results

Participants (age range 24-62 years) included 7 women, 13 daily smokers, and 7 nondaily smokers (Table 2). Thirteen participants were cigarette-only smokers and 7 reported dual use of cigarettes and e-cigarettes. Participants reported an average of 7.3 (SD 3.3) years of residence in the United States. Half of our participants were restaurant staff. A total of 9 main themes and 5 subthemes emerged from the data.

Table 2. Demographic and smoking characteristics of participants (N=20).

Characteristics	Values
Gender, n (%)	
Male	13 (65)
Female	7 (35)
Age, mean (SD) years	37.4 (10.4)
Length of residence in the United States, mean (SD) years	7.3 (3.3)
Education, n (%)	
Middle school or less	4 (20)
High school or vocational high school	10 (50)
Some college, no degree or associate degree	4 (20)
Bachelor's or advanced degree	1 (5)
Unreported	1 (5)
Occupation, n (%)	
Restaurant staff	10 (50)
Hairdresser	5 (25)
Taxi driver	2 (10)
Housewife	2 (10)
Retired	1 (5)
Current smoking status, n (%)	
Nondaily smoker	7 (35)
Daily smoker	13 (65)
Cigarette consumption per day, mean (SD)	9.4 (5.4)
Age of smoking initiation, mean (SD) years	19.8 (5.9)
Current e-cigarette use, n (%)	
Yes	7 (35)
No	13 (65)

Domain 1: Quitting Experience

Theme 1: Quit Attempt and Intention

Nineteen participants reported that they had tried to quit, but only 2 had used evidence-based quitting methods (both had used nicotine patches or gum). Fourteen participants wanted to quit completely, and their reasons were related to concerns for their own health and the health of their children. Of the 14 participants, 4 planned to quit immediately or after finishing the last few packs of cigarettes, 1 planned to quit in 3 to 4 months, 2 stated that they would quit in the future, when they planned to have a baby, and 7 were interested in quitting but had no plans about when to quit.

> *I want to quit completely, including cigarettes and e-cigarettes, because smoking is bad for health. It hurts my lungs. But I haven't thought about when to quit. Probably in the future when something happens to me, I will make the determination to quit.* [Participant #4, male, 27 years old, daily smoker]

Subtheme: Reasons for Lacking Determination in Quitting

Participants who were ambivalent about quitting discussed why they were not determined to quit. Restaurant staff often claimed that smoking was their only excuse for taking a break from work, which had a big impact on their decision. Other factors included perceived low risks of smoking, low confidence in quitting, and need for socializing with friends.

> *I don't see the possibility to quit because my job [as restaurant staff] requires high-intensity labor work. We're busy all the time and I'm exhausted. We have no breaks because the boss doesn't allow us to take a break. I have the excuse because I smoke so I can take a short break. If I quit, I would no longer have an excuse. So I'm not gonna quit unless I change the job.* [Participant #16, male, 37 years old, daily smoker]

> *I don't think I need to quit, because smoking doesn't seem to have an effect on me. It may because I don't smoke much... I have a minor issue with breathing*

but I'm not sure if it's related to smoking. [Participant #14, female, 36 years old, nondaily smoker]

I don't want to quit because I tried [to quit] last time and I couldn't deal with the withdrawal. It was miserable. [Participant #1, male, 29 years old, daily smoker]

Domain 2: Perceptions About the Messages

Theme 2: Likes About Educational Messages

More than half of the participants reported that the messages provided them with information that they were unaware of before, particularly about the adverse health effects of smoking. Some participants were in favor of the content about how smoking damages the body.

Cigarettes contain lots of toxic chemicals which is bad for health. I know it. But I don't know the exact harms caused by smoking or how exactly it damages the body. The messages provide detailed, comprehensive information about the harms, and explain clearly how smoking affects health… I know smoking causes lung cancer, but I don't know it causes stroke and heart disease as well. This is new [knowledge] to me. [Participant #15, male, 39 years old, daily smoker]

Subtheme: Change in Motivation to Quit

Half of the participants reported that, after reading the messages, they would consider quitting (among those who were ambivalent about quitting) or became more motivated to quit (among those who planned to quit). This could be in part because they became more informed about the health impacts of smoking.

The messages about the dangers of smoking really scared me! While reading the messages, I was thinking "oh no, my heart cannot go wrong. I'm still young." I think quitting is important. [Participant #8, female, 36 years old, daily smoker]

Subtheme: Factors That Influenced Participant Decisions to Quit

One message described common reasons why smokers want to quit. When asked to select factors that were important in their decision-making about quitting, participants generally chose health-related items (eg, "be healthier," "look healthier," "improve the health of people around you," and "enjoy better sexual and reproductive health"). Participants rarely selected other items (eg, "save money," "live in a better environment," "live longer," "be free of addiction," and "shape your family and community"). When asked why saving money was not among the top factors that influenced their decisions, nondaily and light smokers noted that the cost of smoking was only a small portion of their daily expenses. Heavy smokers often reported that they had access to cheaper cigarettes through alternative sources.

I smoke a pack every 2 or 3 days so it's like $3 per day. This is a tiny amount of money, just like a cup of coffee or a piece of bread... People do not spend hundreds of dollars to buy cartons of cigarettes at one time. So no one will make the calculation to see how much it [smoking] would cost for a day, for a month, and for a year. [Participant #10, male, 28 years old, daily smoker]

Theme 3: Likes About Self-Efficacy Messages

Many participants reported that the self-efficacy messages offered encouragement. They particularly favored the content emphasizing the role of the environment and quitting skills in successful attempts. Participants expressed strong agreement with the message emphasizing that quitting takes many attempts.

The messages pointed out the importance of environment, which is THE reason why I failed years ago. It tells exactly what I feel… While reading these messages, I felt more confident in myself. [Participant #19, female, 51 years old, daily smoker]

For the majority of smokers, quitting is impossible to achieve with one or two tries. "There is no real failure in the process of quitting." This is a great point. For those who want to quit, knowing this will make them more confident in themselves. [Participant #10, male, 28 years old, daily smoker]

Theme 4: Dislikes About Self-Efficacy Messages

According to some participants, content about the influence of a smoker's success in quitting on his or her community was exaggerated. Participants felt that quitting is "not a big deal" (Participant #17, female, 36 years old, nondaily smoker).

I don't think other people really care whether you quit or not. If you quit, you may feel good about yourself. You may think you're successful. But it's not a big achievement. It's your own business. It cannot inspire others. [Participant #15, male, 39 years old, daily smoker]

Theme 5: Likes About Skill Messages

Participants overwhelmingly favored the skill messages and described them as "useful," "concrete," "comprehensive," and "relevant." Most quitting strategies were perceived to be feasible, including redesigning the environment, identifying smoking triggers, and using substitute behaviors to cope with nicotine cravings. Participants claimed they were willing to apply the strategies in future quit attempts. When exploring appropriate ways to refuse offered cigarettes, participants preferred to say "no" directly, claim that they have quit, or give a reason or an excuse (eg, "I plan to have a baby" or "my throat hurts").

The suggestions are very concrete. … The information is comprehensive, complete, and perfect! As long as you follow the guidance, you will definitely succeed! [Participant #6, male, 56 years old, daily smoker]

Theme 6: Mixed Opinions About Skill Messages

Participants reported mixed opinions toward engaging with their social support networks in the quitting process. Female smokers were generally in favor of social support, and they had all applied this strategy in previous quit attempts. In contrast, male smokers were often reluctant to engage family or friends. Some

stated that "quitting is a personal decision." Some noted that family's criticism about their quitting efforts undermined their determination, which explained their preference against engaging social networks in quit attempts.

> *If I decide to quit, I'll tell my family and friends. I'll ask them to remind me. I think it will be helpful.* [Participant #14, female, 36 years old, nondaily smoker]

> *I will not purposively tell others that I'm quitting. It's your own business.* [Participant #13, male, 29 years old, daily smoker]

> *[If I'm going to quit] I won't tell my family. Dealing with cravings is already very hard. They [my family] are likely to keep nagging and making sarcastic comments [on my slip] and the thing [quitting] will eventually go the wrong way...* [Participant #2, male, 30 years old, daily smoker]

Participants also had different opinions about setting a quit date. Half of the participants favored using Mondays as a quit (or requit) date because (1) "it makes me more want to try [quitting] because even I fail I can always do it again" (Participant #14, female, 36 years old, nondaily smoker), (2) "The date is fixed. It's easy to remember" (Participant #9, male, 62 years old, daily smoker), and (3) "I can see my progress... The first week I may stay abstinent [from Monday] till Tuesday. Next week I may stay abstinent till Wednesday" (Participant #15, male, 39 years old, daily smoker). Other participants were skeptical about the effect of setting a quit date or preferred to have their own quit date.

> *No one can really quit from the quit date so there's no need to set a date. If you decide to quit, do it now!* [Participant #5, male, 27 years old, daily smoker]

Subtheme: Challenges in Applying the Quitting Strategies

Participants, particularly restaurant staff, discussed challenges in redesigning the environment. They reported that it was nearly impossible to create a supportive environment because most coworkers smoked. Participants also noted the challenge of using substitute behaviors to cope with craving and stress. For taxi drivers and restaurant staff, their stress from work was often a trigger to smoke, but they did not know how to deal with it.

> *I work in the kitchen. All the cooks smoke, so it's impossible to have a good environment to quit.* [Participant #19, female, 51 years old, daily smoker]

> *I only smoke during the working days. The kitchen work is very busy. … Every few hours, I tell the manager that "I need to smoke." So I can have a break and relax for a few minutes.* [Participant #17, female, 36 years old, nondaily smoker]

Theme 7: Readability, Format, Length, and Language Clarity

Participants were generally able to summarize the messages appropriately and captured the key content. They expressed a preference for messages written in bulleted lists and limited to approximately 380 Chinese characters (including spaces). That length fits within about a single screen of a mobile phone (in the default font). Participants were reluctant to read long paragraphs and some would skip those long messages.

Participants were not familiar with some terms, including "quit date," "social smoking," and "substitute behavior." How we translated the terms into Chinese influenced their understanding of the concepts.

> *I've never heard of "quit date." Does it mean you cannot start quitting until the "quit date"?* [Participant #13, male, 29 years old, daily smoker]

Domain 3: Attitudes Toward the WeChat Mobile Messaging Smoking Cessation Intervention

Theme 8: Attitudes Toward the WeChat Cessation Intervention

Participants expressed strong interest in the WeChat smoking cessation intervention. They felt that the intervention would provide them with useful information, take a minimum of their time, and not require them to quit at a certain date or take medications.

> *I'd like to participate because it [the intervention] does not force me to do anything, like no smoking starting from a certain day... It's like a daily reminder. It may help me reduce smoking.* [Participant #1, male, 29 years old, daily smoker]

Subtheme: Privacy Concerns

The majority of participants reported no concerns about joining the WeChat mobile messaging cessation intervention, except 2 smokers who noted concerns about privacy issues associated with using WeChat.

Theme 9: Suggestions on the Frequency and Timing of Messages

All the participants reported that receiving 1 daily message was acceptable. Some noted that "2 or 3 messages per day" or "no more than 5 messages per day" would be appropriate. Preferences for the timing of the messages varied, ranging from morning ("8:00 AM when I go to work") to late night ("10:00 PM when I finish a whole day work"). Some participants had no preference for timing, because they had no or minimal access to their mobile phone while at work or they checked messages only when they wanted to.

> *I usually check [WeChat] messages when I wake up in the morning and after work… While at work, we put phones away at the cashier. We get phones back after work or if we need to make an emergent call.* [Participant #17, female, 36 years old, nondaily smoker]

Discussion

Principal Findings

This study describes the development of a WeChat-based mobile messaging program for smoking cessation tailored to Chinese immigrant smokers. Our findings provide important insights into how smokers perceived the messages and their attitudes

toward the WeChat mobile messaging cessation intervention. In general, the messages were well-received. Participants described the messages as useful and relevant, and claimed that the content enhanced their motivation to quit and offered encouragement. Participants particularly liked messages about the harms of smoking and strategies for quitting (eg, coping skills), and reported that they learned new information from these messages. This finding affirms the low health literacy for smoking and lack of quitting skills among Chinese immigrant smokers. Previous studies have reported that Chinese immigrant smokers are often unclear about the exact health effects of smoking, and that many smokers inaccurately believe that quitting would result in health problems or that willpower is the only key to cessation [12,13,15-17,40]. Our findings indicate that smoking cessation programs that work with Chinese immigrants have to include concrete information about the specific dangers of tobacco use, the health benefits of quitting, and detailed guidance on how to quit.

A noteworthy finding is that the factor that most impacted participant decision-making on quitting was concern for their health and the health of their children. Other factors, such as economic gains, did not seem to have an equal effect in motivating smokers to quit. Among smokers who planned to quit, all cited health concerns as the reason for quitting. Findings suggest that cessation programs for Chinese immigrant smokers need to emphasize the health impacts of smoking, rather than stress the financial impacts.

Participants expressed a strong interest in the WeChat mobile messaging cessation intervention for several reasons. First, they felt the messages were helpful. Second, they thought that the intervention would take a minimum of time. Thus, they perceived no barriers to engagement. First-generation immigrants often report long and inflexible working hours, which prevents them from participating in quitlines or in-person cessation treatment programs [12-14,17]. Mobile messaging intervention has the potential to accommodate immigrant smokers, because they can access messages at their own pace and at the time of their convenience. Leveraging the WeChat platform made the cessation intervention more accessible. Third, unlike most SMS cessation programs that solely focus on smokers who are committed to quitting (with a self-determined quit date in the next 30 days) [25-27], our intervention was designed to target a larger audience, including smokers not ready to quit. This appealed to our participants, who claimed that they would give it a try because the intervention "does not force me to quit." Our next step involves testing the feasibility, acceptability, and effectiveness of the WeChat mobile messaging cessation intervention among Chinese immigrant smokers.

We identified needs for message modifications. For example, our message library should include more content about how to cope with stress and offer tips on how to stay abstinent at work. Several studies have reported that stress is a primary trigger for smoking among Chinese immigrant smokers and that the lack of coping skills is an important factor associated with the low intention to quit [12,39]. In addition, smoking is more prevalent among workers in certain occupations (eg, food preparation and services) [42]. Smokers working in these industries may encounter more difficulties in quitting. Hence, an emphasis on capacity building, including how to manage stress and deal with workplace smoking, should be integrated into smoking cessation programs that focus on immigrant smokers. We also identified a preference for the message format used in our WeChat mobile messaging intervention. Paragraphs should be written in bulleted lists when possible. Long paragraphs are not appropriate, since they are hard to read on mobile phones.

Limitations

The interviews were conducted in one geographic area with Chinese immigrant smokers who were already somewhat interested in quitting. The findings may not be generalizable to other areas or to smokers who are not at all interested in quitting. Moreover, the interviews were all conducted in Mandarin. We lack data and perspectives on the messages and attitudes toward the WeChat mobile messaging cessation intervention from smokers who speak Cantonese or other dialects. However, Chinese immigrants in New York City are mostly able to communicate in Mandarin, even if they primarily speak Cantonese [43].

Conclusions

This study contributes to the small body of literature on the development process of mobile messaging smoking cessation interventions tailored to specific racial or ethnic minority groups. The participants in our study claimed that the messages enhanced their motivation to quit and offered encouragement. Their report on barriers to quitting guided us in modifying the messages and generating new ones. The participants were strongly interested in the WeChat mobile messaging cessation intervention and noted its potential to address their access barriers to existing smoking cessation programs. We will follow up on this qualitative study with feasibility testing and a randomized controlled trial to explore if a WeChat mobile messaging cessation intervention is acceptable, and to what extent it can help Chinese immigrant smokers achieve abstinence.

Acknowledgments

This study was supported by the National Institute on Minority Health and Health Disparities at the National Institutes of Health (K01MD014165) and New York University Grossman School of Medicine's Department of Population Health (Global Health Pilot Grant Funding Award). JCL was supported by the National Cancer Institute at the National Institutes of Health (F32CA265059). The funding organizations played no role in study design, data collection and analysis, manuscript writing, or the decision to submit the paper for publication. The authors are grateful for Ms. Le Li's assistance with data collection and coding.

Conflicts of Interest

LH reports holding stock in Tencent Holdings Limited. The authors have no other conflicts to declare.

References

1. U.S. Department of Health and Human Services. The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health; 2014.
2. Epi Data Brief. City of New York. 2019. URL: https://www1.nyc.gov/assets/doh/downloads/pdf/epi/databrief103.pdf [accessed 2022-06-16]
3. Sansone N, Yong HH, Li L, Jiang Y, Fong GT. Perceived acceptability of female smoking in China. Tob Control 2015 Nov;24 Suppl 4:iv48-iv54 [FREE Full text] [doi: 10.1136/tobaccocontrol-2015-052380] [Medline: 26240136]
4. Chen MJ, Tang H. Review of smoking cessation research among Asian Americans: the state of the research. Nicotine Tob Res 2007 Sep;9 Suppl 3:485-493. [doi: 10.1080/14622200701587052] [Medline: 17978977]
5. Liu JJ, Wabnitz C, Davidson E, Bhopal RS, White M, Johnson MR, et al. Smoking cessation interventions for ethnic minority groups--a systematic review of adapted interventions. Prev Med 2013 Dec;57(6):765-775. [doi: 10.1016/j.ypmed.2013.09.014] [Medline: 24076130]
6. FitzGerald JM, Poureslami I, Shum J. Assessing beliefs and risk perceptions on smoking and smoking cessation in immigrant Chinese adult smokers residing in Vancouver, Canada: a cross-sectional study. BMJ Open 2015 Feb 03;5(2):e006435 [FREE Full text] [doi: 10.1136/bmjopen-2014-006435] [Medline: 25649211]
7. Zhu SH, Cummins SE, Wong S, Gamst AC, Tedeschi GJ, Reyes-Nocon J. The effects of a multilingual telephone quitline for Asian smokers: a randomized controlled trial. J Natl Cancer Inst 2012 Feb 22;104(4):299-310 [FREE Full text] [doi: 10.1093/jnci/djr530] [Medline: 22282542]
8. Wu D, Ma GX, Zhou K, Zhou D, Liu A, Poon AN. The effect of a culturally tailored smoking cessation for Chinese American smokers. Nicotine Tob Res 2009 Dec;11(12):1448-1457 [FREE Full text] [doi: 10.1093/ntr/ntp159] [Medline: 19915080]
9. Ma GX, Fang C, Shive SE, Su X, Toubbeh JI, Miller S, et al. A culturally enhanced smoking cessation study among Chinese and Korean smokers. International Electronic Journal of Health Education 2005;8:1-10 [FREE Full text]
10. Tsoh JY, Burke NJ, Gildengorin G, Wong C, Le K, Nguyen A, et al. A Social Network Family-Focused Intervention to Promote Smoking Cessation in Chinese and Vietnamese American Male Smokers: A Feasibility Study. Nicotine Tob Res 2015 Aug;17(8):1029-1038 [FREE Full text] [doi: 10.1093/ntr/ntv088] [Medline: 26180229]
11. Chen C, Anderson CM, Babb SD, Frank R, Wong S, Kuiper NM, et al. Evaluation of the Asian Smokers' Quitline: A Centralized Service for a Dispersed Population. Am J Prev Med 2021 Mar;60(3 Suppl 2):S154-S162 [FREE Full text] [doi: 10.1016/j.amepre.2020.01.033] [Medline: 33663703]
12. Jiang N, Zhang Y, Qian X, Thorpe L, Trinh-Shevrin C, Shelley D. Chinese immigrant smokers' access barriers to tobacco cessation services and experience using social media and text messaging. Tob Prev Cessat 2020;6:52 [FREE Full text] [doi: 10.18332/tpc/125942] [Medline: 33083680]
13. Katigbak C, Maglalang DD, Chao YY, Au H, Liang W, Zuo S. Cultural Perspectives on Tobacco Use and Cessation Among Chinese American Immigrants: A Community-Engaged Qualitative Study. J Transcult Nurs 2019 Jul;30(4):350-358. [doi: 10.1177/1043659618817583] [Medline: 30556472]
14. Spence W, Zhu L. Perceptions of smoking cessation among Glasgow's Chinese community. Tob Prev Cessat 2017;3:127 [FREE Full text] [doi: 10.18332/tpc/77942] [Medline: 32432201]
15. Tsang IK, Tsoh JY, Wong C, Le K, Cheng JW, Nguyen AN, et al. Understanding and use of nicotine replacement therapy and nonpharmacologic smoking cessation strategies among Chinese and Vietnamese smokers and their families. Prev Chronic Dis 2014 Feb 20;11:E26 [FREE Full text] [doi: 10.5888/pcd11.130299] [Medline: 24556252]
16. Saw A, Paterniti D, Fung L, Tsoh JY, Chen MJ, K Tong E. Social Environmental Influences on Smoking and Cessation: Qualitative Perspectives Among Chinese-Speaking Smokers and Nonsmokers in California. J Immigr Minor Health 2017 Dec;19(6):1404-1411 [FREE Full text] [doi: 10.1007/s10903-016-0358-6] [Medline: 26872641]
17. Mao A, Bottorff JL. A Qualitative Study on Unassisted Smoking Cessation Among Chinese Canadian Immigrants. Am J Mens Health 2017 Nov;11(6):1703-1712 [FREE Full text] [doi: 10.1177/1557988315627140] [Medline: 26819181]
18. Fu SS, Ma GX, Tu XM, Siu PT, Metlay JP. Cigarette smoking among Chinese Americans and the influence of linguistic acculturation. Nicotine Tob Res 2003 Dec;5(6):803-811. [doi: 10.1080/14622200310001614566] [Medline: 14668064]
19. Ma GX, Tan Y, Toubbeh J, Su X. Differences in stages of change of smoking behavior among current smokers of four Asian American subgroups. Addict Behav 2003 Oct;28(8):1431-1439. [doi: 10.1016/s0306-4603(03)00071-6] [Medline: 14512065]
20. Shelley D, Fahs M, Scheinmann R, Swain S, Qu J, Burton D. Acculturation and tobacco use among Chinese Americans. Am J Public Health 2004 Feb;94(2):300-307. [doi: 10.2105/ajph.94.2.300] [Medline: 14759946]
21. Yu ESH, Chen EH, Kim KK, Abdulrahim S. Smoking among Chinese Americans: behavior, knowledge, and beliefs. Am J Public Health 2002 Jun;92(6):1007-1012. [doi: 10.2105/ajph.92.6.1007] [Medline: 12036797]

22. Spigner C, Yip MP, Huang B, Tu SP. Chinese and Vietnamese adult male smokers' perspectives regarding facilitators of tobacco cessation behavior. Asian Pac J Cancer Prev 2007;8(3):429-435 [FREE Full text] [Medline: 18159982]
23. Zhang JY, Chan SS, Fong DY, Malone RE, Lam TH. The social context of smoking cessation in China: an exploratory interview study. Tob Control 2012 Jan;21(1):57-58. [doi: 10.1136/tc.2010.040691] [Medline: 21459932]
24. Kong G, Ells DM, Camenga DR, Krishnan-Sarin S. Text messaging-based smoking cessation intervention: a narrative review. Addict Behav 2014 May;39(5):907-917 [FREE Full text] [doi: 10.1016/j.addbeh.2013.11.024] [Medline: 24462528]
25. Scott-Sheldon LAJ, Lantini R, Jennings EG, Thind H, Rosen RK, Salmoirago-Blotcher E, et al. Text Messaging-Based Interventions for Smoking Cessation: A Systematic Review and Meta-Analysis. JMIR Mhealth Uhealth 2016 May 20;4(2):e49 [FREE Full text] [doi: 10.2196/mhealth.5436] [Medline: 27207211]
26. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y, Dobson R. Mobile phone text messaging and app-based interventions for smoking cessation. Cochrane Database Syst Rev 2019 Oct 22;10:CD006611 [FREE Full text] [doi: 10.1002/14651858.CD006611.pub5] [Medline: 31638271]
27. Smokefree Text Messaging Programs. Smokefree.gov. URL: https://smokefree.gov/tools-tips/text-programs [accessed 2019-10-11]
28. Most popular global mobile messenger apps as of July 2021, based on number of monthly active users (in millions). Statista. URL: https://www.statista.com/statistics/258749/most-popular-global-mobile-messenger-apps/ [accessed 2021-10-02]
29. Abroms LC, Whittaker R, Free C, Mendel Van Alstyne J, Schindler-Ruwisch JM. Developing and Pretesting a Text Messaging Program for Health Behavior Change: Recommended Steps. JMIR Mhealth Uhealth 2015 Dec 21;3(4):e107 [FREE Full text] [doi: 10.2196/mhealth.4917] [Medline: 26690917]
30. Fiore M, Jaén CR, Baker TB, Benowitz NL, Curry SJ, Dorfman SF, et al. Treating Tobacco Use and Dependence: 2008 Update. Rockville, MD: U.S. Department of Health and Human Services, Public Health Service; May 2008.
31. Abroms LC, Boal AL, Simmens SJ, Mendel JA, Windsor RA. A randomized trial of Text2Quit: a text messaging program for smoking cessation. Am J Prev Med 2014 Sep;47(3):242-250 [FREE Full text] [doi: 10.1016/j.amepre.2014.04.010] [Medline: 24913220]
32. Heminger CL, Boal AL, Zumer M, Abroms LC. Text2Quit: an analysis of participant engagement in the mobile smoking cessation program. Am J Drug Alcohol Abuse 2016 Jul;42(4):450-458. [doi: 10.3109/00952990.2016.1149591] [Medline: 27120396]
33. Haug S, Schaub MP, Venzin V, Meyer C, John U. Efficacy of a text message-based smoking cessation intervention for young people: a cluster randomized controlled trial. J Med Internet Res 2013 Aug 16;15(8):e171 [FREE Full text] [doi: 10.2196/jmir.2636] [Medline: 23956024]
34. Sutton S, Smith S, Jamison J, Boase S, Mason D, Prevost AT, et al. Study protocol for iQuit in Practice: a randomised controlled trial to assess the feasibility, acceptability and effectiveness of tailored web- and text-based facilitation of smoking cessation in primary care. BMC Public Health 2013 Apr 10;13:324 [FREE Full text] [doi: 10.1186/1471-2458-13-324] [Medline: 23575031]
35. Naughton F, Cooper S, Foster K, Emery J, Leonardi-Bee J, Sutton S, et al. Large multi-centre pilot randomized controlled trial testing a low-cost, tailored, self-help smoking cessation text message intervention for pregnant smokers (MiQuit). Addiction 2017 Jul;112(7):1238-1249 [FREE Full text] [doi: 10.1111/add.13802] [Medline: 28239919]
36. Jiang N, Nguyen N, Siman N, Cleland CM, Nguyen T, Doan HT, et al. Adaptation and Assessment of a Text Messaging Smoking Cessation Intervention in Vietnam: Pilot Randomized Controlled Trial. JMIR Mhealth Uhealth 2021 Oct 08;9(10):e27478 [FREE Full text] [doi: 10.2196/27478] [Medline: 34623318]
37. Cartujano-Barrera F, Arana-Chicas E, Catley D, Cox LS, Diaz FJ, Ellerbeck EF, et al. Decídetexto: Mobile cessation support for Latino smokers. Study protocol for a randomized clinical trial. Contemp Clin Trials 2020 Dec;99:106188 [FREE Full text] [doi: 10.1016/j.cct.2020.106188] [Medline: 33080379]
38. Cartujano-Barrera F, Peña-Vargas CI, Arana-Chicas E, Pérez-Ramos JG, Mattei J, Hurtado-de-Mendoza A, et al. Decídetexto: Feasibility and Acceptability of a Mobile Smoking Cessation Intervention in Puerto Rico. Int J Environ Res Public Health 2021 Feb 03;18(4):1-12 [FREE Full text] [doi: 10.3390/ijerph18041379] [Medline: 33546156]
39. Poureslami I, Shum J, Aran N, Tregobov N. Chinese- and English-speaking adult current smokers' perspectives on smoking and culturally and linguistically appropriate cessation: a qualitative analysis. Addict Sci Clin Pract 2020 Jul 06;15(1):23 [FREE Full text] [doi: 10.1186/s13722-020-00197-4] [Medline: 32631420]
40. Tu SP, Walsh M, Tseng B, Thompson B. Tobacco Use by Chinese American Men: An Exploratory Study of the Factors Associated with Cigarette Use and Smoking Cessation. Asian Am Pac Isl J Health 2000;8(1):46-57 [FREE Full text] [Medline: 11567512]
41. Bingham A, Witkowsky P. Deductive and Inductive Approaches to Qualitative Data Analysis. In: Vanover C, Mihas P, Saldana J, editors. Analyzing and Interpreting Qualitative Research: After the Interview. New York, NY: SAGE Publications; 2021.
42. Syamlal G, Mazurek JM, Hendricks SA, Jamal A. Cigarette smoking trends among U.S. working adult by industry and occupation: findings from the 2004-2012 National Health Interview Survey. Nicotine Tob Res 2015 May;17(5):599-606 [FREE Full text] [doi: 10.1093/ntr/ntu185] [Medline: 25239956]
43. Semple K. In Chinatown, Sound of the Future is Mandarin. New York Times 2009 Oct 22:A1 [FREE Full text]

Abbreviations

SCT: social cognitive theory
SMS: short message service

Edited by A Mavragani; submitted 31.12.21; peer-reviewed by L Struik, N Nguyen; comments to author 06.04.22; revised version received 01.06.22; accepted 08.06.22; published 30.06.22.

Please cite as:
Jiang N, Rogers ES, Cupertino P, Zhao X, Cartujano-Barrera F, Lyu JC, Hu L, Sherman SE
Development of a WeChat-based Mobile Messaging Smoking Cessation Intervention for Chinese Immigrant Smokers: Qualitative Interview Study
JMIR Form Res 2022;6(6):e36091
URL: https://formative.jmir.org/2022/6/e36091
doi:10.2196/36091
PMID:35771603

Original Paper

The Association Between the Use of Low-Slice Computed Tomography Machines and Downstream Care: Comparative Study of 16-Slice and 64-Slice Computed Tomography Angiography

Adam C Powell[1], PhD; James W Long[2], BSBA; Uday U Deshmukh[1], MPH, MD; Jeffrey D Simmons[2], MPH, MD

[1]HealthHelp, Houston, TX, United States

[2]Humana Inc, Louisville, KY, United States

Corresponding Author:
Adam C Powell, PhD
HealthHelp
16945 Northchase Drive, Suite 1300
Houston, TX, 77060
United States
Phone: 1 6179399168
Email: powellad@healthhelp.com

Abstract

Background: Although computed tomography (CT) studies on machines with more slices have reported higher positive and negative predictive values, the impact of using low-slice (16-slice) CT machines on downstream testing has not been well studied. In community outpatient settings, low-slice CT machines remain in use, although many hospitals have adopted higher-slice machines.

Objective: This study examines the association between the use of low-slice CT machines and downstream invasive testing in the context of the CT angiography of the neck.

Methods: Included health insurance claims pertained to adults with commercial or Medicare Advantage health plans who underwent the CT angiography of the neck. Site certification data were used to assign counts of slices to claims. Claims that were made in the 60 days after CT were examined for cervicocerebral angiography. The association between the number of slices and cervicocerebral angiography was evaluated by using a chi-square test and multivariate logistic regression.

Results: Claims for 16-slice CT had a 5.1% (33/641) downstream cervicocerebral angiography rate, while claims for 64-slice CT had a 3.1% (35/1125) rate, and a significant difference (P=.03) was observed. An analysis that was adjusted for patient demographics also found a significant relationship (odds ratio 1.64, 95% CI 1.00-2.69; P=.047).

Conclusions: The use of low-slice CT machines in the community may impact the quality of care and result in more downstream testing.

(JMIR Form Res 2022;6(6):e32892) doi:10.2196/32892

KEYWORDS

computed tomography; tomography; diagnostic imaging; outpatient; angiography; obsolescence; computed tomography angiography of the neck; neck; low-slice computed tomography; cervicocerebral angiography; downstream testing; computed tomography machine; invasive testing; machine; testing; invasive

Introduction

Although there have been great advances in computed tomography (CT) technology, when undergoing outpatient imaging in the community, many patients continue to have imaging performed at underresourced facilities with a single, often low-slice CT machine. Further, while high-slice CT machines with 128 detector rows (slices) or more are often available at academic medical centers, in the community, 16-slice CT (low-slice CT) machines remain in active use, and the use of 64-slice CT (medium-slice CT) machines is common. This study aims to explore whether the use of low-slice CT at some facilities has the potential to impact patient outcomes.

As CT machines have advanced, systems with increasing numbers of detector rows have become available. In 2004, the first clinical images from a 64-slice CT machine were released

to the public, and a machine was commercially launched in the American market [1]. In part due to the increased costs of 64-slice CT machines, their adoption has not been universal. By 2016—over 1 decade after 64-slice CT became available—only 63% of hospitals had access to a CT machine with 64 or more slices [2]. Thus, a substantial number of patients continue to undergo CT on machines with fewer slices. Outpatient facilities with a single CT machine may not have medium-slice or high-slice machines, as they may lack the volumes necessary for justifying their purchase. Nonetheless, the use of low-slice machines may impact patients' care, and facilities that are only able to offer 16-slice CT may wish to consider upgrading if the use of such technology has an impact on downstream care.

The benefits of 64-slice CT, relative to 16-slice CT, have primarily been examined in terms of the quality of visualization, positive predictive value, negative predictive value, sensitivity, and specificity [3-6]. However, researchers have not yet examined the association between the use of 64-slice and 16-slice CT machines and downstream testing. This study provides preliminary insights into the association between the number of slices and downstream testing in a narrow context to explore whether an association may exist. It has historically been difficult to study how the type of CT machine used impacts downstream testing within the context of a broad population because health insurance claims data—the main source of information for this type of study—do not provide any information about the CT machine used to acquire an image. Although electronic medical records may contain relevant information, these data may be challenging to use when answering this question because patients may undergo testing at multiple facilities, thereby creating issues with linking CT scans to downstream testing. Furthermore, health care providers may not have ample data related to both 16-slice and 64-slice CT, as providers typically own a small number of CT machines (if they have more than 1 CT machine), thereby hampering comparisons between the two types of machines.

The better the information that can be gathered from a CT machine, the greater the extent to which it may be a substitute for other forms of testing. Current evidence suggests that as CT technology advances, the potential to replace coronary angiography for the evaluation of coronary artery disease with CT scans increases [7]. If higher-quality CT imaging can reduce the use of angiography, it has the potential to improve the welfare of patients, as angiography can result in complications. A review of 19,826 patients who underwent diagnostic cerebral angiography found that neurological complications occurred in 2.63% of patients, with 0.14% experiencing strokes resulting in permanent disability [8].

In order to assess the potential benefits of the use of CT machines with a greater number of slices in underresourced outpatient settings, this study examined the association between the use of 16-slice and 64-slice CT machines to perform the CT angiography of the neck and subsequent cervicocerebral angiography within a population of patients that had not recently undergone head/neck imaging, catheterization, or percutaneous coronary interventions and had undergone imaging at facilities with only a single type of CT machine. If patients who underwent CT via a 64-slice machine were less likely to undergo subsequent cervicocerebral angiography than those who underwent CT via a 16-slice machine, then there may be quality benefits to 64-slice CT, in addition to the visual and informational benefits that have previously been characterized [3-6]. If the use of low-slice CT machines impacts the quality of care that patients receive, facilities may need to consider upgrading their equipment, and ordering physicians may need to more closely consider the capabilities of the equipment that is present at the imaging facilities to which they make referrals. Although low-slice CT machines remain in use in the community, used medium-slice and high-slice CT machines are readily available on the secondary market, and many facilities offer superior forms of CT.

Methods

Data Source and Sample Population

All health insurance claims for the CT angiography of the neck (current procedural terminology code: 70498) with dates of service ranging between September 15, 2017, and September 14, 2018, were extracted from the database of a national health care organization. Health insurance claims pertained to adults with commercial and Medicare Advantage health plans. The dates of the CT scans served as the index dates. Claims were excluded from this study if they pertained to patients who were not continuously enrolled in their health plan from 90 days prior to the index date to 60 days following the index date. To restrict this study to patients beginning new episodes of care, claims were likewise excluded if they pertained to patients who had received a CT image, magnetic resonance image, or positron emission tomography image of the head or neck in the 90 days prior to the index date or if they pertained to patients who had undergone catheterization or a percutaneous coronary intervention in the same time period. Claims could only be linked to a CT machine with a known number of slices if CT was performed with a CT machine that was based at a facility participating in an outpatient site of a service certification program and if the site of service had either 1 CT machine or multiple CT machines with the same number of slices. Claims were excluded from this study if they could not be matched to a CT machine with a known number of slices. Finally, claims were excluded if they pertained to a CT machine with a number of slices other than 16 or 64.

Ethics Approval

This study was reviewed and approved by Advarra's institutional review board (approval number: Pro00033618). The institutional review board granted a waiver of informed consent for this study due to its aggregate, observational nature. This study was conducted in accordance with the Declaration of Helsinki.

Measurement

The dependent variable in the analysis was whether a CT claim was followed by a claim for cervicocerebral angiography (current procedural terminology codes: 36221-36228) within 60 days. The independent variable was whether a CT image was acquired via a 16-slice CT machine or a 64-slice CT machine. The covariates that were included as potential

confounders in the analysis were variables for whether the patients pertaining to the claims lived in areas with above or below the median income of the sample, as well as their sex, urbanicity, health plan's line of business (commercial vs Medicare), and age (<65 years vs ≥65 years). The average income within a patient's zip code was determined by using the median income of the past 12 months in 2017 inflation–adjusted US dollars, as reported by the American Community Survey [9]. Rural areas were identified by using a zip code mapping table that was developed by the Centers for Medicare & Medicaid Services [10].

Analysis

Descriptive statistics were calculated for the sample. Chi-square tests were conducted to assess whether the population that underwent 16-slice CT differed from the population that underwent 64-slice CT on each of the potential confounders. Chi-square tests were also conducted to test for a univariate association between each of the variables within the study and the outcome—whether a patient underwent downstream cervicocerebral angiography. A multivariate logistic regression was conducted to determine the adjusted association between the number of slices that the used CT machine possessed and downstream cervicocerebral angiography. The results from the multivariate logistic regression were reported as odds ratios. Throughout the analysis, a *P* value of <.05 was used as the threshold for determining statistical significance.

Results

As shown in Figure 1, of the 41,063 claims that met the initial inclusion criteria, 1766 remained qualified for this study after the consideration of the exclusion criteria. Many of the exclusions occurred due to situations in which it was infeasible to match claims to specific CT machines. The claims occurred at 252 different rendering facilities, of which 121 had 16-slice CT machines and 131 had 64-slice CT machines. Descriptive statistics are presented in Table 1. Among the total cohort, 36.3% (641/1766) of claims were related to CT that was performed with a 16-slice machine, while the majority were related to CT that was performed with a 64-slice machine (1125/1766, 64.7%). The patients who underwent CT had a mean age of 71 years and were from communities with a mean local income of US $57,460. Most of the claims pertained to patients who had Medicare Advantage plans (1620/1766, 91.7%); only 8.3% (146/1766) pertained to patients who had commercial insurance. There was a significant association between community income and the number of slices in the CT machine used (*P*<.001); 46.3% (297/641) of claims for a 16-slice CT came from a community with below the median income, while 37.2% (418/1125) of claims for a 64-slice CT came from a community with below the median income.

As shown in Table 2, chi-square tests found that none of the control variables had a significant univariate association with downstream cervicocerebral angiography. However, a chi-square test found that there was a statistically significant univariate relationship between the performance of CT via a 16-slice machine or 64-slice machine and subsequent cervicocerebral angiography (*P*=.03). Claims for 16-slice CT had a 5.1% (33/641) subsequent cervicocerebral angiography rate, while claims for 64-slice CT had a 3.1% (35/1125) subsequent cervicocerebral angiography rate (Figure 2).

The adjusted analysis, which is shown in Table 3, found that there was a significant association between the performance of CT via a 16-slice machine or 64-slice machine and subsequent cervicocerebral angiography (*P*=.047; odds ratio 1.64, 95% CI 1.00-2.69). None of the control variables in the adjusted analysis had a significant or near-significant association with cervicocerebral angiography.

Figure 1. Sample selection diagram. CT: computed tomography; MRI: magnetic resonance imaging; PCI: percutaneous coronary intervention; PET: positron emission tomography.

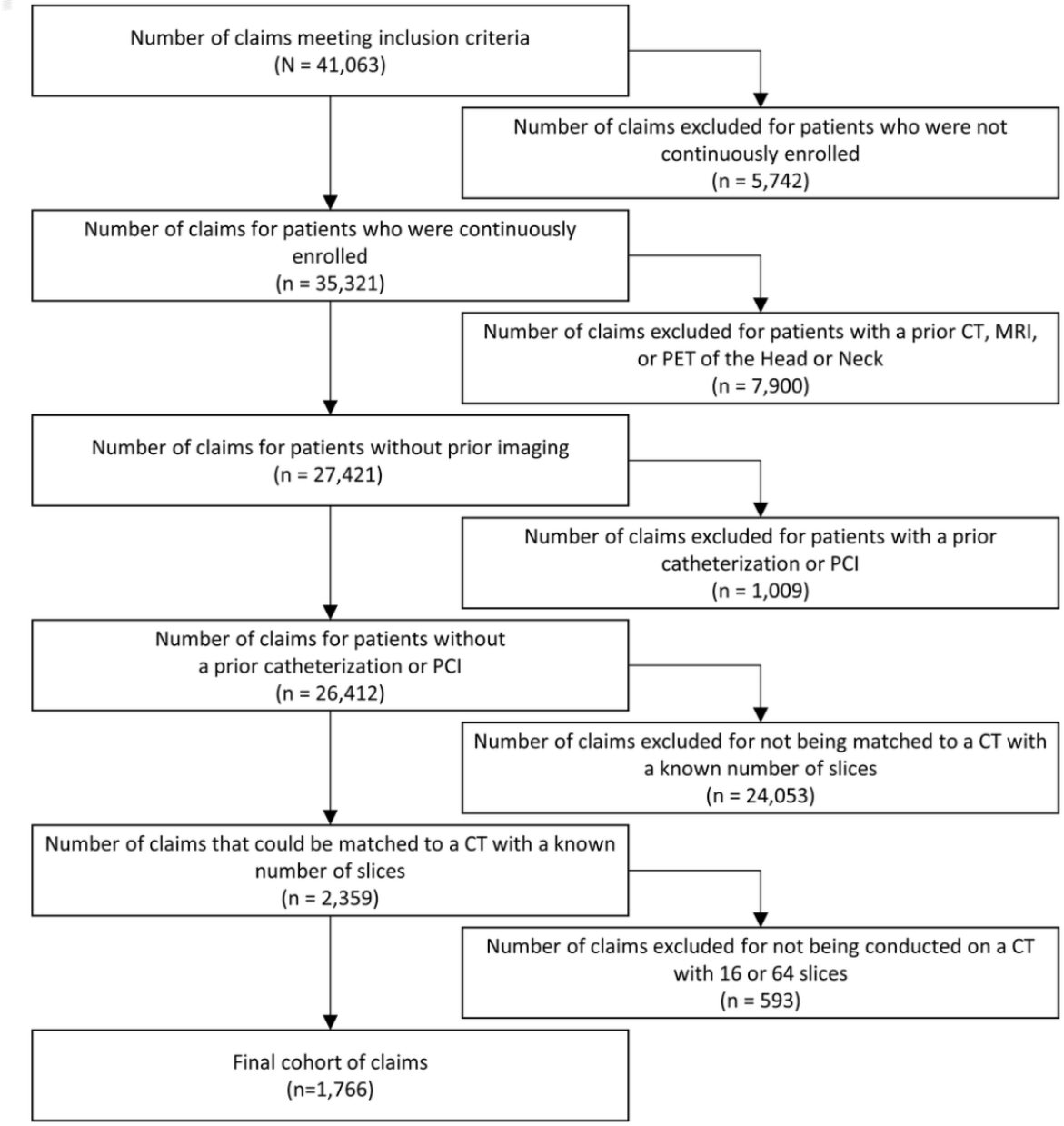


Table 1. Descriptive statistics.

Patient characteristic	All claims (N=1766), n (%)	Claims for 16-slice CT[a] (n=641), n (%)	Claims for 64-slice CT (n=1125), n (%)	*P* value
Below median income (vs above median income or omitted)	715 (40.5)	297 (46.3)	418 (37.2)	<.001
Male (vs female)	844 (47.8)	310 (48.4)	534 (47.5)	.72
Rural (vs urban)	260 (14.7)	93 (14.5)	167 (14.8)	.85
Commercial (vs Medicare)	146 (8.3)	34 (5.3)	112 (10)	.001
Aged under 65 years (vs ≥65 years)	304 (17.2)	96 (15)	208 (18.5)	.06

[a]CT: computed tomography.

Table 2. Univariate associations between variables and downstream cervicocerebral angiography.

Patient characteristic	All claims (N=1766), n (%)	Claims not pertaining to cervicocerebral angiography (n=1698), n (%)	Claims pertaining to cervicocerebral angiography (n=68), n (%)	*P* value
Below median income (vs above median income or omitted)	715 (40.5)	685 (40.3)	30 (44.1)	.53
Male (vs female)	844 (47.8)	812 (47.8)	32 (47.1)	.90
Rural (vs urban)	260 (14.7)	252 (14.8)	8 (11.8)	.48
Commercial (vs Medicare)	146 (8.3)	142 (8.4)	4 (5.9)	.47
Aged under 65 years (vs ≥65 years)	304 (17.2)	296 (17.4)	8 (11.8)	.23
16-slice CT[a] (vs 64-slice CT)	641 (36.3)	608 (35.8)	33 (48.5)	.03

[a]CT: computed tomography.

Figure 2. Visual depiction of the downstream cervicocerebral angiography rate following 16-slice CT or 64-slice CT. Each cell represents 1 claim for CT. Shaded cells represent the proportion of CTs that were followed by downstream cervicocerebral angiography. CT: computed tomography.

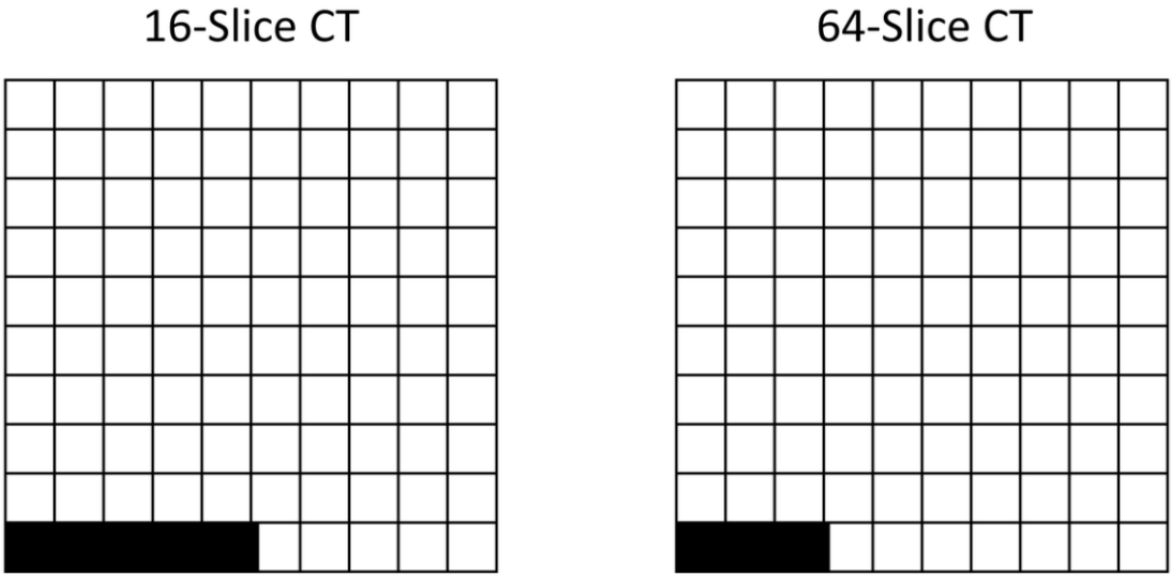


Table 3. Adjusted odds ratios from the multivariate logistic regression.

Patient characteristic	Odds ratio	95% CI	*P* value
Below median income (vs above median income)	1.17	0.70-1.93	.55
Male (vs female)	0.97	0.60-1.59	.92
Rural (vs urban)	0.71	0.30-1.45	.38
Commercial (vs Medicare)	1.05	0.27-3.37	.94
Aged under 65 (vs ≥65 years)	0.63	0.23-1.42	.31
16-slice CT[a] (vs 64-slice CT)	1.64	1.00-2.69	.047

[a]CT: computed tomography.

Discussion

Principal Findings

Compared to patients who had their neck CT angiography performed with a 16-slice machine, patients who had their neck CT angiography performed with a 64-slice machine were significantly (*P*=.03) less likely to undergo a subsequent cervicocerebral angiography. If the patients who were referred for 16-slice CT were not believed a priori to be more difficult to conclusively diagnose than those referred to 64-slice CT, the findings suggest that the use of 64-slice CT machines may improve the quality of care by reducing the need for downstream invasive testing. Although they are not universally available, 64-slice CT machines are widely distributed and may be the best choice for patients requiring a CT scan. Considering the data presented herein, physicians may want to review the

specifications of the CT machines that a rendering facility possesses when referring their patients.

The findings from this study pertain to care that was delivered in 2017 and 2018—over 1 decade after the introduction of 64-slice CT. Of the 2359 total CT claims that could be tied to a CT machine with a known number of slices (Figure 1), 641 claims were tied to a 16-slice CT (Table 1), representing 27.2% of the total. Although 64-slice CT was performed more frequently than 16-slice CT in the context examined, the findings suggest that 16-slice CT machines remain in active use in many facilities.

The findings of this study are congruent with those of prior research on the benefits of performing CT by using the greatest number of slices available. A previous study found that when a 64-slice CT machine was used to assess left ventricular function, the values it produced were more similar to those obtained via echocardiography and technetium-99m gated single-photon emission CT compared to the values produced by a 16-slice CT machine [11]. A multicenter prospective study on the utility of 16-slice versus 64-slice CT in screening patients for coronary artery disease did not detect significant differences between 16-slice and 64-slice CT in terms of sensitivity, specificity, positive predictive value, or negative predictive value when CT-based findings were compared to coronary angiography–based findings, although it is possible that the study was underpowered [5]. Furthermore, a literature review of the diagnostic performance of 16-slice versus 64-slice CT, in comparison with coronary angiography, found that 64-slice CT had higher sensitivity values, specificity values, positive predictive values, and negative predictive values [4].

Finally, although community income was not a significant (P=.55) variable in the adjusted model for estimating the determinants of cervicocerebral angiography, community income has an association with the number of slices that the used CT machine possessed. There may be an access disparity issue wherein patients from lower-income communities are less able to access 64-slice CT. Although evidence of a clinical impact resulting from this was found by this study, the association between income and the nature of the CT machine used may require future investigation.

In our study, we examined 1 current procedural terminology code for neck CT, which represented only a small fraction of the overall CT imaging conducted in the United States. In 2017, America's 39 million traditional Medicare beneficiaries collectively underwent 16 million CTs, which were billed through their Medicare Part B benefits [12,13]. Millions of other CTs were performed on patients with commercial, Medicaid, and Medicare Advantage health plans. Although not all examinations may benefit from being performed via 64-slice CT rather than 16-slice CT, it is possible that there are other indications for which the benefit of 64-slice CT could be demonstrated. Further research on the downstream consequences of the choice of a CT machine has the potential to impact many people due to the large number of CTs that are conducted in the United States each year.

Limitations

There are several limitations that need to be considered when interpreting our findings. As this was a claims-based analysis, it is unknown whether patients who underwent 16-slice CT or 64-slice CT differed in terms of factors other than those that were examined. It is possible that clinical differences influenced assignment. For the findings to have been a product of biased assignment, physicians would need to have preferred assigning patients with more ambiguous cases to 16-slice CT, which is a choice that seems counterintuitive.

The findings of this study may not be representative of the care that is delivered to the overall population, as only outpatient facilities with a single CT machine or a set of CT machines with the same number of slices could be included in the analysis. Thus, the sample does not include facilities with a diverse set of CT machines. This requirement forced the exclusion of many of the available claims. If ordering physicians' choice of a rendering facility is influenced by proximity rather than by the nature of their CT machine, then their choice might serve to counterbalance the potential for assignment bias.

Lastly, the population studied was not representative of the overall population of the United States. The individuals included in this study resided predominantly in the south, as this is where the health care organization that supplied the data had the strongest presence. The sample likewise did not contain anyone with traditional Medicare or anyone with a Medicaid plan lacking dual eligibility for Medicare. As the incomes of the patients in the sample are unknown, the average incomes within their zip codes were used as a proxy. Given the average age of the patients in the sample, it is likely that many were retired and were earning incomes lower than those that were typical for their communities.

Conclusions

The analysis found a significant association between the performance of CT via a 16-slice CT machine or 64-slice CT machine and subsequent cervicocerebral angiography before (P=.03) and after (P=.047) adjusting for patient demographic factors. None of the other factors examined had a significant association with subsequent cervicocerebral angiography. When patients can potentially have access to imaging via 64-slice CT, ordering physicians should consider the potential benefits of directing patients to undergo the CT angiography of the neck via a 64-slice CT machine rather than a 16-slice CT machine. Further research is needed to explore the impact of the decision to use low-slice CT machines in additional clinical contexts and examine whether the relationship remains significant after controlling for clinical factors. Although higher-slice CT machines may not be readily available in some communities, our findings suggest that physicians need to weigh the benefits of access against the benefits of having patients undergo a diagnostic examination that is less likely to result in subsequent downstream testing.

Conflicts of Interest

The authors report the following competing interests. ACP and UUD report an employment or consulting relationship with HealthHelp/WNS at the time this paper was written. JWL and JDS report employment by Humana Inc. ACP additionally reports employment by Payer+Provider Syndicate; stock ownership of Berkshire Hathaway, Community Health Systems, CVS Health Corp, Hospital Corporation of America Healthcare, Payer+Provider Syndicate, Quorum Health Corp, and Tenet Healthcare Corp; and research support from the Max Institute of Healthcare Management. UUD additionally reports stock ownership of Johnson & Johnson, Merck, Pfizer, Express Scripts, Halyard Health Inc, Cigna, Proctor & Gamble, and WNS Holdings, as well as honoraria for participation in the Start Time Optimization of Biologics in Polyarticular Juvenile Idiopathic Arthritis Advisory Panel.

References

1. Siemens releases first clinical images from 64-slice CT scanner. J Clin Eng 2004;29(3):125-126. [doi: 10.1097/00004669-200407000-00032]
2. Percentage of hospitals with access to multi-slice spiral computed tomography: 64 + slice. Harvey L. Neiman Health Policy Institute. URL: https://www.neimanhpi.org/data_series/percentage-of-hospitals-with-access-to-multi-slice-spiral-computed-tomography-64-slice-2/#/0 [accessed 2020-10-08]
3. Seifarth H, Ozgün M, Raupach R, Flohr T, Heindel W, Fischbach R, et al. 64- versus 16-slice CT angiography for coronary artery stent assessment: in vitro experience. Invest Radiol 2006 Jan;41(1):22-27. [doi: 10.1097/01.rli.0000191365.88345.e1] [Medline: 16355036]
4. Hamon M, Morello R, Riddell JW, Hamon M. Coronary arteries: diagnostic performance of 16- versus 64-section spiral CT compared with invasive coronary angiography--meta-analysis. Radiology 2007 Dec;245(3):720-731. [doi: 10.1148/radiol.2453061899] [Medline: 17951354]
5. Marano R, De Cobelli F, Floriani I, Becker C, Herzog C, Centonze M, NIMISCAD Study Group. Italian multicenter, prospective study to evaluate the negative predictive value of 16- and 64-slice MDCT imaging in patients scheduled for coronary angiography (NIMISCAD-Non Invasive Multicenter Italian Study for Coronary Artery Disease). Eur Radiol 2009 May;19(5):1114-1123. [doi: 10.1007/s00330-008-1239-8] [Medline: 19089430]
6. Baumüller S, Leschka S, Desbiolles L, Stolzmann P, Scheffel H, Seifert B, et al. Dual-source versus 64-section CT coronary angiography at lower heart rates: comparison of accuracy and radiation dose. Radiology 2009 Oct;253(1):56-64. [doi: 10.1148/radiol.2531090065] [Medline: 19587311]
7. Khan R, Rawal S, Eisenberg MJ. Transitioning from 16-slice to 64-slice multidetector computed tomography for the assessment of coronary artery disease: are we really making progress? Can J Cardiol 2009 Sep;25(9):533-542 [FREE Full text] [doi: 10.1016/s0828-282x(09)70144-8] [Medline: 19746244]
8. Kaufmann TJ, Huston J3, Mandrekar JN, Schleck CD, Thielen KR, Kallmes DF. Complications of diagnostic cerebral angiography: evaluation of 19,826 consecutive patients. Radiology 2007 Jun;243(3):812-819. [doi: 10.1148/radiol.2433060536] [Medline: 17517935]
9. S1903: Median income in the past 12 months (in 2017 inflation-adjusted dollars). United States Census Bureau. URL: https://data.census.gov/cedsci/table?q=S1903&tid=ACSST5Y2017.S1903 [accessed 2021-05-02]
10. DME-Rural-Zip-and-Formats. Centers for Medicare and Medicaid Services. 2016. URL: https://tinyurl.com/2p9ae3fv [accessed 2021-05-06]
11. Abbara S, Chow BJW, Pena AJ, Cury RC, Hoffmann U, Nieman K, et al. Assessment of left ventricular function with 16- and 64-slice multi-detector computed tomography. Eur J Radiol 2008 Sep;67(3):481-486. [doi: 10.1016/j.ejrad.2007.07.022] [Medline: 17826936]
12. Computed tomography procedures per 1000 beneficiaries (Medicare Part B by state of residence). Harvey L. Neiman Health Policy Institute. URL: https://tinyurl.com/45pnzwa4 [accessed 2022-05-31]
13. MDCR ENROLL AB 10. Original Medicare enrollment: Part A and/or Part B enrollees, by age group, calendar years 2012-2017. Centers for Medicare & Medicaid Services. URL: https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/CMSProgramStatistics/2017/Downloads/MDCR_ENROLL_AB/2017_CPS_MDCR_ENROLL_AB_10.pdf [accessed 2021-05-05]

Abbreviations

CT: computed tomography

Edited by G Eysenbach; submitted 13.08.21; peer-reviewed by N Maglaveras, G Nneji; comments to author 01.05.22; revised version received 10.05.22; accepted 18.05.22; published 30.06.22.

Please cite as:
Powell AC, Long JW, Deshmukh UU, Simmons JD
The Association Between the Use of Low-Slice Computed Tomography Machines and Downstream Care: Comparative Study of 16-Slice and 64-Slice Computed Tomography Angiography
JMIR Form Res 2022;6(6):e32892
URL: https://formative.jmir.org/2022/6/e32892
doi:10.2196/32892
PMID:35771601

Original Paper

Finding Primary Care—Repurposing Physician Registration Data to Generate a Regionally Accurate List of Primary Care Clinics: Development and Validation of an Open-Source Algorithm

Ian R Cooper[1], MSc; Cameron Lindsay[1], MSc; Keaton Fraser[1]; Tiffany T Hill[1], MA; Andrew Siu[1], MBA; Sarah Fletcher[1], MSc, PhD; Jan Klimas[1,2], PhD; Michee-Ana Hamilton[1], MSc; Amanda D Frazer[1], MSc; Elka Humphrys[1], PhD; Kira Koepke[1], MSc; Lindsay Hedden[3,4], PhD; Morgan Price[1], MD, PhD, CCFP; Rita K McCracken[1], MD, PhD

[1]Innovation Support Unit, Department of Family Practice, University of British Columbia, Vancouver, BC, Canada

[2]British Columbia Centre on Substance Use, Vancouver, BC, Canada

[3]Faculty of Health Sciences, Simon Fraser University, Burnaby, BC, Canada

[4]British Columbia Academic Health Sciences Network, Vancouver, BC, Canada

Corresponding Author:
Rita K McCracken, MD, PhD
Innovation Support Unit
Department of Family Practice
University of British Columbia
3rd Floor David Strangway Building
Vancouver, BC, V6T 1Z3
Canada
Phone: 1 604 827 4168
Email: rita.mccracken@ubc.ca

Abstract

Background: Some Canadians have limited access to longitudinal primary care, despite its known advantages for population health. Current initiatives to transform primary care aim to increase access to team-based primary care clinics. However, many regions lack a reliable method to enumerate clinics, limiting estimates of clinical capacity and ongoing access gaps. A region-based complete clinic list is needed to effectively describe clinic characteristics and to compare primary care outcomes at the clinic level.

Objective: The objective of this study is to show how publicly available data sources, including the provincial physician license registry, can be used to generate a verifiable, region-wide list of primary care clinics in British Columbia, Canada, using a process named the Clinic List Algorithm (CLA).

Methods: The CLA has 10 steps: (1) collect data sets, (2) develop clinic inclusion and exclusion criteria, (3) process data sets, (4) consolidate data sets, (5) transform from list of physicians to initial list of clinics, (6) add additional metadata, (7) create working lists, (8) verify working lists, (9) consolidate working lists, and (10) adjust processing steps based on learnings.

Results: The College of Physicians and Surgeons of British Columbia Registry contained 13,726 physicians, at 2915 unique addresses, 6942 (50.58%) of whom were family physicians (FPs) licensed to practice in British Columbia. The CLA identified 1239 addresses where primary care was delivered by 4262 (61.39%) FPs. Of the included addresses, 84.50% (n=1047) were in urban locations, and there was a median of 2 (IQR 2-4, range 1-23) FPs at each unique address.

Conclusions: The CLA provides a region-wide description of primary care clinics that improves on simple counts of primary care providers or self-report lists. It identifies the number and location of primary care clinics and excludes primary care providers who are likely not providing community-based primary care. Such information may be useful for estimates of capacity of primary care, as well as for policy planning and research in regions engaged in primary care evaluation or transformation.

(JMIR Form Res 2022;6(6):e34141) doi:10.2196/34141

KEYWORDS

physicians, primary care; primary health care; health services accessibility; practice patterns, physicians; physicians' offices; computing methodologies; algorithms

Introduction

Improving access to primary care is on health agendas around the world [1]. This is likely linked to the finding that increasing supply of primary care physicians is associated with decreased mortality rates [2]. In Canada, primary care accessibility is a persistent challenge. The media regularly note a supposed family physician (FP) shortage [3] despite most provinces having the highest ever number of FPs per capita [4].

Primary care is the first point of access to the health care system and provides longitudinal, person-focused care for most care needs across the lifespan [5,6]. The majority of primary care in Canada is still delivered by FPs at community-based outpatient clinics [7]. Currently, systematically identifying clinics versus individual physicians is difficult. This may be due to the ongoing reliance on funding of the majority of primary care services via individual physician remuneration [7-9] versus using a centralized system of service delivery, as is more commonly seen with other social services, such as public schools.

Across North America, initiatives to transform primary care have addressed access to care by establishing primary care teams [7-11]. Coordinated, team-based care in a primary care clinic is recognized as a critical part of modernized care [12-15] and has begun to be seen in Canada. However, analyses of these transformations often still rely on using the individual physician as a unit of service delivery [16,17] despite indicators that FPs may work in multiple locations, in a combination of roles [18-20], and that clinic culture and organization may contribute to physicians' behaviors that influence quality of care [21,22]. The primary care clinic, rather than the individual physician, is evolving as the main access point for many patients. Exclusive reliance on physician-centric metrics potentially fails to include the contribution of nonphysician team members to the accessibility and quality of patient care. Other regions have begun to frame descriptions of primary care features and outcomes using the clinic as the unit of analysis [8,23,24]. A comprehensive list of primary care clinic locations is necessary for the effective assessment of initiatives that aim to improve access and would provide a baseline from which to measure change. There is no complete list available in British Columbia; an environmental scan has provided a few partial lists of specific types of care provision (eg, locations funded by health authorities) and some local data sources relying on physician self-report and voluntary listing of clinics. Other health care regions face similar data challenges as well as a need to improve access to primary care [25-27].

The objective of this study is to develop and verify an algorithm that uses a continually updated, public listing of individual physician license registration addresses to create an accurate region-wide listing of clinic locations where primary care is delivered.

Methods

Overview

This study describes the development, application, and verification of an algorithm applied to a physician license registration list in order to reliably generate an accurate list of primary care clinics. The Clinic List Algorithm (CLA) was developed by the primary care Innovation Support Unit at the University of British Columbia as part of a larger project examining primary care capacity and access measurement.

Study Setting

British Columbia, Canada, has a single-payer health system [7], and primary care is provided almost exclusively by FPs. In many regions, physicians must register with a licensing body (eg, provincial or territorial college of physicians and surgeons), declare their specialization qualification (eg, family medicine), and give an address at which they provide services [28]. These registrations may also include other useful information, such as date of graduation, medical school location, or additional demographic data. In Canada, registration lists are publicly available.

Data Sources

We used the following publicly available data sources to create the CLA.

The Registry

The College of Physicians and Surgeons of British Columbia (CPSBC) physician registry is the base to which we will apply additional data sets, processing actions, and verifying actions, and it is referred to in this paper as the Registry. It is updated continuously, is publicly accessible online, and can be requested from the CPSBC in a more accessible format [29]. The version accessed for this study is from September 2020.

Additional Address Data

The BC Ministry of Health publishes a comprehensive list of regional health authorities, broken down into Community Health Service Areas (CHSAs) [30]. We incorporated CHSAs because related health profiles exist detailing a community's demographic, socioeconomic, and health and disease status. DataBC geolocation services information was used to add longitude and latitude coordinates to addresses, in accordance with their terms of service in British Columbia [31].

Partial Lists of Specific Family Physician Workplaces

Walk-in Clinic, Urgent and Primary Care Centre, and Hospital Lists

These lists were accessed between September 2020 and March 2021. The BC Ministry of Health publishes a list of locations in British Columbia that "provide walk-in treatment services for people who have minor illnesses or injuries or injuries that do not require a visit to a hospital emergency department or an urgent care facility" [32]. Walk-in treatment services are a form of community-based primary care [33]. This walk-in clinic list is updated twice yearly, in December and June. The Ministry of Health also manages two other publicly available lists that identify all hospitals [34] and Urgent and Primary Care Centres (UPCCs) [35] in British Columbia. These are updated similarly to the walk-in clinic list. UPCCs have been available in British Columbia since 2019 and "provide a flexible resource to meet the urgent and primary health care need" in British Columbia [35].

Corrections Facilities List

We generated a list of corrections facilities in British Columbia from multiple sources of publicly available information. These facilities included federal and provincial institutions, as well as immigration and remand institutions. At these addresses, primary care services are delivered to people living in corrections facilities. This care is not accessible to members of the community.

Long-term Care Facilities List

We obtained a file of all registered long-term care facilities from the BC Office of the Seniors Advocate [36]. This registry is regularly updated and contains details for all BC long-term care homes in which there are publicly funded beds. Medical services provided in these locations require a patient to meet strict admission criteria and are not available to other community members.

Variables

The outcome of interest was an accurate list of primary care clinics. A clinic was defined as a location where primary care services are delivered by at least one FP registered to practice in British Columbia. Clinic locations were derived from the addresses provided by physicians who registered for a license to practice. The CLA was used to translate the physician registration addresses to primary care clinic locations (see Results section).

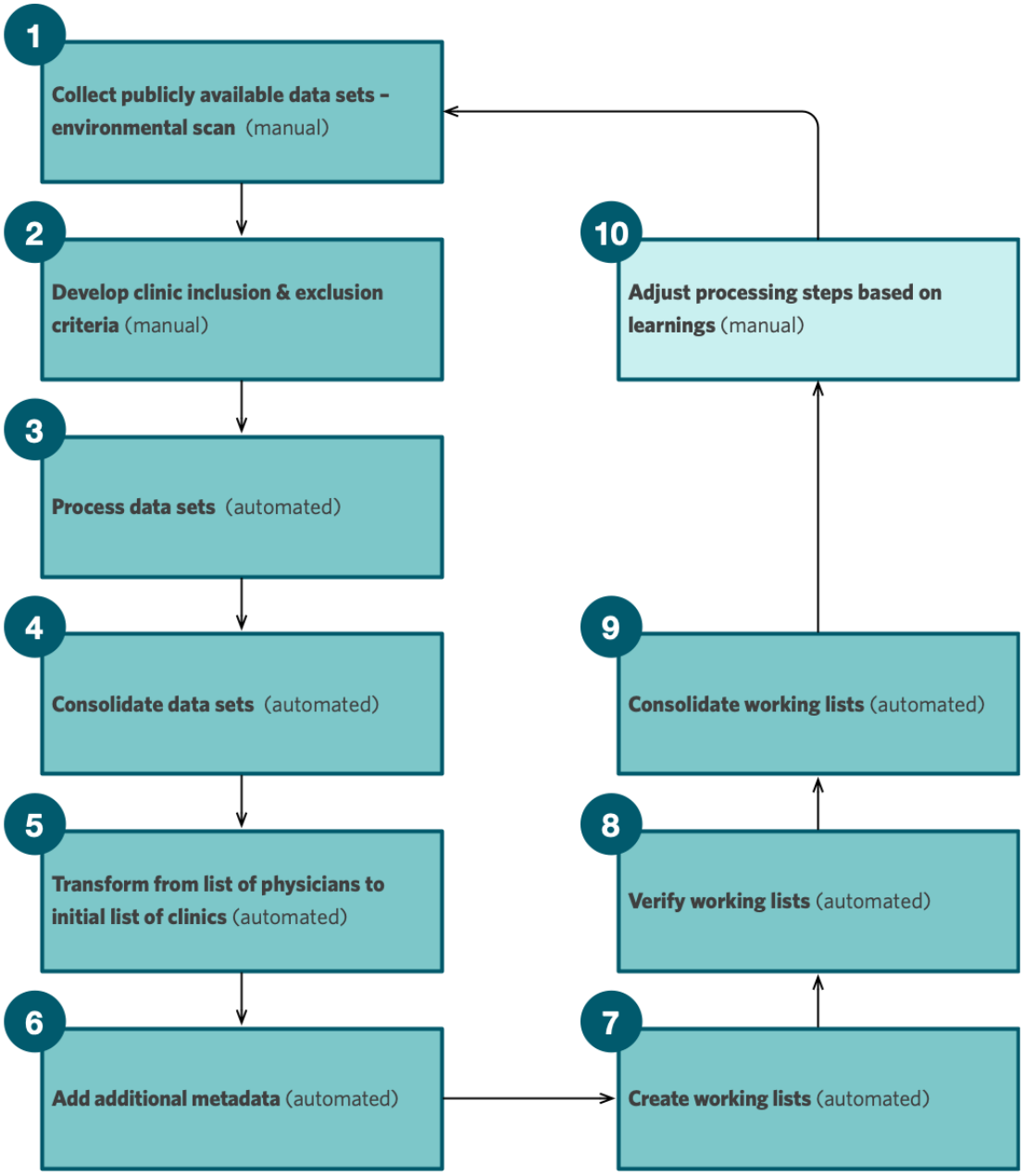
Key Informants

Members of the research team with intimate knowledge of the regional primary care system (RKM and IRC) were needed in helping to develop inclusion and exclusion criteria.

Analytic Methods

The complete software code used for this algorithm is published online on GitHub [37]. The software was developed using Python (version 3.7.7; Python Software Foundation) in conjunction with the following open-source libraries: NumPy, pandas, OpenPyXL, GeoPy, and RegEx. Figure 1 summarizes the steps and flow developed for the algorithm application process, and Multimedia Appendix 1 includes additional details about how each step was completed. The Results section describes the application of the CLA. Descriptive statistics were calculated using Microsoft Excel for Mac (version 16.61).

Figure 1. Clinic List Algorithm steps to generate the primary care location list using publicly available data sources.



Ethics Approval

This study is exempt from Research Ethics Board review, as the data used for this research falls under the criteria outlined in Article 2.2 of the Tri-Council Policy Statement.

Results

Step 1: Collect Publicly Available Data Sets—Environmental Scan (Manual)

The data sources were collected and collated, including the Registry and additional address data. We plan to create regional service maps, and this level of location detail will facilitate that future work. Some addresses, almost exclusively post-office boxes, had to be reviewed individually and altered manually to allow them to be processed through the BC Address Geocoder. We then collected the following partial lists as sources for inclusion or exclusion: the walk-in clinic, UPCC, and hospital lists, which were accessed between September 2020 and March 2021, were generated; the corrections facilities list was generated; and the long-term care facilities list was requested.

Step 2: Develop Clinic Inclusion and Exclusion Criteria (Manual)

We established inclusion and exclusion criteria for the type 2 secondary data sources. Walk-in clinics and UPCCs were included. Hospitals, corrections facilities, and long-term care facilities were excluded. Key search terms were developed from author knowledge (RKM and IRC) and grey and published literature searches. Search terms are shown in Table 1.

Table 1. Search terms used in filtering and sorting addresses from the Registry list of the College of Physicians and Surgeons of British Columbia to each of our working lists.

Criteria and working lists[a]	Search terms[b]
Inclusion criteria	
Walk-in clinic	No search terms: list developed from walk-in clinic list
Urgent and primary care	No search terms: list developed from Urgent and Primary Care Centre list
Family	Family and (Med* or Clinic or Centre or Center or Associate* or Care or Practice)
First Nations	First Nation* or First People* or Native or Aboriginal or Indigenous or {clinic-specific name}
Clinic or center	Clinic* or Associate* or Center or Centre of Practice or Doctor*
Exclusion criteria	
Hospital	Hosp*: list developed from hospital list
Long-term care	Lodging or Lodge* or Manor or Senior or ALC[c]: list developed from long-term care list
Corrections	Correction* or Immigra* or Pretrial or Custody or Institution or Detention or Holding or Healing Village: list developed from corrections list
Sexual health	Sexual or STI[d] or STD[e] or {clinic specific name}
Women's health	Wom[a,e]n* or Menopause or Matern* or Birth* or Obstetric* or Gyne*
Virtual	{organization-specific name} or Virtual or E[-]Health or Tele* or I[-]Health
Administrative	Airport or Consulting or Admin* or Fraser Health Authority or First Nation* Health Authority or Coroner or CPSBC[f] or College of Physician* and Surgeon* or Health Canada or VCH[g] or Worksafe or Worksafebc or Worker* Comp* or BCAA[h] or Veteran* Affair or RCMP[i] or Air Canada or Quality

[a]Working lists created in step 7 of the Clinic List Algorithm.

[b]For search terms, the asterisk (*) denotes that any character following the search term was allowed, round brackets signify the inclusion of any number of terms inside those brackets, square brackets indicate that any of the letters inside the brackets were allowable in the word, and curly brackets identify a search term specific to a clinic that is not a generic term and is not included in this list. Regular expression statements used in the algorithm code are found in Multimedia Appendix 2.

[c]ALC: alternate level of care.

[d]STI: sexually transmitted infection.

[e]STD: sexually transmitted disease.

[f]CPSBC: College of Physicians and Surgeons of British Columbia.

[g]VCH: Vancouver Coastal Health.

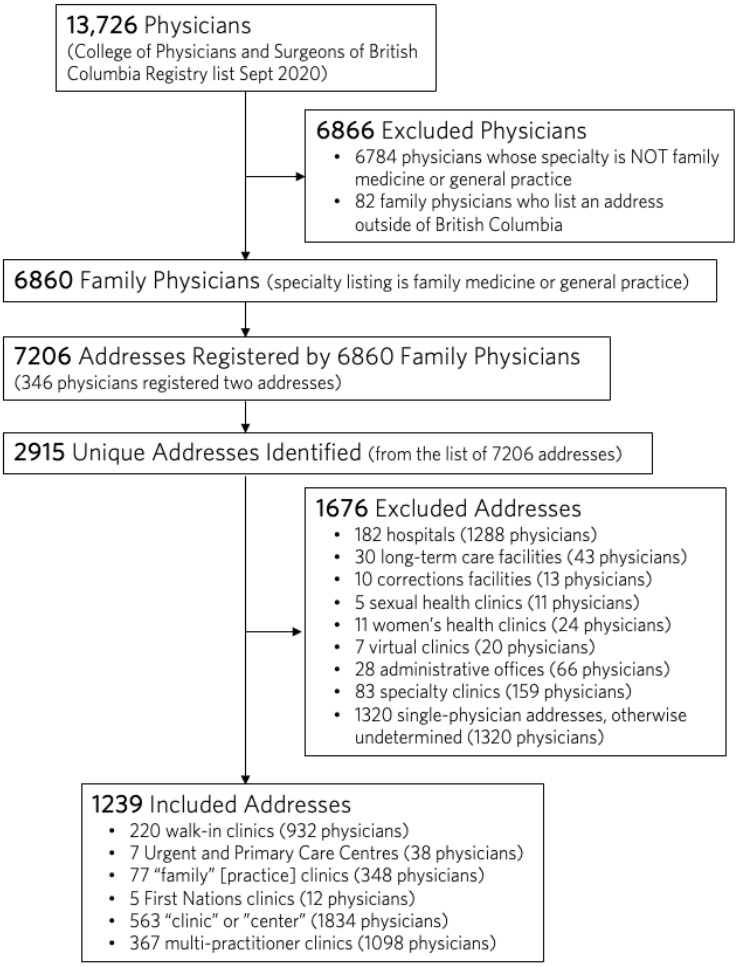
[h]BCAA: British Columbia Automobile Association.

[i]RCMP: Royal Canadian Mounted Police.

Step 3: Process Data Sets (Automated)

The software filtered the CPSBC Registry list using the Specialties and Certificates column to include only FPs. The CPSBC Registry list contained 13,726 physicians, of which 6942 were identified as FPs. We removed addresses from outside British Columbia (n=82), leaving 6860 FPs (Figure 2). These 6860 physicians were found to have 7206 registered addresses. All remaining address fields in the Registry list were then standardized.

Figure 2. Case study results identifying included, excluded, and undetermined addresses of primary care locations in British Columbia.



Step 4: Consolidate Data Sets (Automated)

Additional address data were merged with the Registry list, appending geolocation coordinates and relevant CHSA data fields to each entry using consistent data fields between data sources.

Step 5: Transform From List of Physicians to Initial List of Clinics (Automated)

The software identified 2915 unique addresses from the Registry list. All 6860 FPs were then assigned to the unique addresses, creating the initial clinic-centric list.

Step 6: Add Additional Metadata (Automated)

Metadata were added as new variables by the software to include the number of FPs identified at each clinic, the rurality of clinic locations, key term identification, and overall summaries of each process undertaken by the software.

Step 7: Create Working Lists (Automated)

The inclusion and exclusion criteria developed in step 2 were applied to create the 15 working lists. The results of this step are depicted in Figure 2. The partial lists' data sources were applied to identify addresses of walk-in clinics, UPCCs, hospitals, corrections facilities, and long-term care facilities and then sorted to their respective working lists. Key search terms in the address fields of each remaining listing were sorted to respective working lists (Table 1). Finally, unique addresses with more than one associated FP and that were not sorted by any previous step were assigned to the multi-practitioner working list. Single addresses listed by only one FP were assigned to the single-practitioner working list.

Step 8: Verify Working Lists (Manual)

The verification process was completed manually, using the working lists produced in step 7. The goal of verification was to determine if an address assigned to a particular working list had been correctly included or excluded as a community-based location that provided any primary care.

The sample size for verification was determined for each separate working list. The proportion that was verified differed between working lists. Considerations in verification were (1) baseline confidence that the sorting method would accurately include and exclude addresses and (2) time and resources available for the verification process. Multimedia Appendix 3 includes the proportion of addresses verified for each group.

The verification process was applied to each working list and consisted of at least one of four reviews. In the first review, lists prepared by government organizations were assumed to be accurate, and no additional verification was done. The second review was manual and was performed by an expert informant (RKM) who personally knew providers or clinics that were delivering primary care. The third and fourth reviews were done by KF and IRC and began with an internet search for clinic

name, address, and FP names. If the internet search was insufficient to ascertain that primary care was being delivered at that location, a final, most resource-intensive level of review was done by phoning the clinic.

The internet search process and script used for phone call verification is included in Multimedia Appendix 4. A team member (IRC) collated responses from all four types of reviews and compared a subset of completed entries to ensure consistency between the methods of verification.

The denominator for verification accuracy was the number of addresses in the step multiplied by the proportion to be verified. If a verified address was included—or excluded, depending on the specific working list—as a primary care location, it contributed to the numerator. This gave a relative accuracy for the groups used in the algorithm to identify inclusion or exclusion of potential primary care clinics.

Confirmation of primary care service provision was not possible for the virtual working list, as the six physical addresses correctly identified with this filter were for administrative offices only. For the remainder of the working lists, verification was completed for almost all potential clinics found, with two notable exceptions. The family working list verification was completed on only 27% (21/77) of the clinics identified because this filter was established late in our development process as a very high-fidelity filter. The single-practitioner working list verification was completed on only 15.98% (211/1320) of the addresses identified; of these 211 addresses, only 118 (55.9%) were currently providing primary care. Initial work to develop patterns to identify clinics from the single-practitioner working list indicated that this working list was nonspecific to primary care clinics (eg, contained many residential addresses).

Step 9: Consolidate Working Lists (Automated)

The software merged the working lists into one final clinic list that is presented in Multimedia Appendix 5. From the original 2915 unique addresses identified, the CLA excluded 356 (12.21%) addresses, representing 1624 FPs. The CLA also defined 1320 addresses as single-physician addresses that were not identified through any other filter. The CLA identified 1239 potential primary care clinics representing 4262 FPs (Figure 2). Table 2 shows the geographic and descriptive data about included primary care clinics.

Table 2. Geographic distribution and descriptive statistics of included primary care clinic addresses from the case study in British Columbia.

CHSA[a] urban or rural classification[b]	Unique addresses[c] (n=1239), n (%)	FPs[d,e] with addresses in this region (n=4262), n (%)	FPs per unique address	
			Median (IQR)	Range
Metropolitan	566 (45.68)	1897 (44.51)	2 (2-4)	1-20
Large urban	189 (15.25)	647 (15.18)	3 (2-4)	1-14
Medium urban	195 (15.74)	717 (16.82)	3 (2-5)	1-23
Small urban	99 (7.99)	410 (9.62)	3 (2-5)	1-21
Rural hub	75 (6.05)	269 (6.31)	3 (2-5)	1-15
Rural	102 (8.23)	299 (7.02)	2 (2-3)	1-15
Remote	12 (0.97)	23 (0.54)	2 (1-2)	1-5
Total	1239 (100)	4262 (100)	2 (2-4)	1-23

[a]CHSA: Community Health Service Area.

[b]Urban and rural classifications are based on the classification of the CHSA by the BC Ministry of Health [30].

[c]The software from the Clinic List Algorithm (CLA) was used to identify the unique addresses that were included, with FPs from the College of Physicians and Surgeons of British Columbia list.

[d]FP: family physician.

[e]These are FPs who are registered with the CPSBC and have listed an address identified by the CLA as a primary care location.

Step 10: Adjust Processing Steps Based on Learnings (Manual)

Each iteration of the working lists was reviewed by the study authors (IRC, CL, KF, TTH, SF, JK, MAH, and RKM). Three revisions of the software were made to address important learnings (eg, adding the term “family” as an inclusion search term and obtaining the complete list of long-term care addresses). The new rules and processes created as a result of discoveries in the verification process were incorporated into the final code, which is found on GitHub [37].

Discussion

Principal Findings

The CLA identified 1239 addresses where primary care was delivered by 4262 FPs from a list of the 6942 FPs licensed to practice in British Columbia. The algorithm used publicly available data sources and could be repeated as the data sources are updated. The CLA was developed to facilitate the study of regional effects of policies to transform primary care, using the clinic as the unit of analysis. This may be of particular value in regions where such lists do not exist or are not publicly available.

Previous work exploring geographic distribution of primary care services also used physician registration addresses to find locations of primary care service delivery [25,26]. However, that work did not include a way to remove the physicians who are unlikely to be providing community-based services. Modern FPs have dynamic and varied practice patterns that rarely are in a single location and providing only community-based primary care [20,38]. The CLA allows for observations about how FPs are organizing their work. We identified 1624 FPs who have an address at which primary care services are not available. While these physicians are likely engaged in important health services work that uses their valuable time and skills, they should not be assumed to be a potential source for primary care [18,38,39].

In our final clinic list, 94.25% (4017/4262) of included FPs shared a registered address with at least one other person. While this may not reflect the complete practice context (eg, whether the practice is organized as a team) and does not include the undetermined single-practitioner working list (n=1320 FPs), it does confirm that most FPs are not working in solo practice settings. The CLA allows for the creation of a unique physician identifier that can be associated with an address. Appropriately constructed research and quality initiatives could use such a variable to study physician practice patterns, identified in administrative data, as mapped to physical locations [8,14,17,40]. In many Canadian regions, this linkage has not previously been possible [5,19,41,42].

The majority of primary care evaluation and research on access and quality of care has needed to rely on the individual FP as the unit of analysis [19,41,43-45] (eg, patient attachment, third next available appointment, and rates of opioid prescribing by a single FP). This does not appear to represent how the majority of FPs are actually practicing. Measurements that focus on the clinic as the unit of analysis for primary care provision would likely better reflect the changing practice context in primary care and allow for accurate health resource planning [8,14,40,46].

Limitations and Future Work

We cannot assume that each practitioner works at each location full time, providing only primary care services, because FPs often work in multiple locations [38]. Further work is required to understand patient access and attachment capacity at each clinic location [47,48]. We assumed that a physician registers all addresses where they provide any care at the time of their license registration. It is possible that this is not the case and that a physician registers an address for convenience or a personal reason. This could result in undercounting of physicians working at a location or failure to identify a primary care location at all.

We neither called nor visited every address to verify their services, due to limited time and resources. We verified only 15.98% (n=211) of the addresses on the single-practitioner working list of 1320 unique addresses held by one FP. Future projects could include a more inclusive verification process for the single-practitioner working list. However, this work would need to be completed manually, given the lack of patterns we were able to discover in software development with the CLA. Future versions of the CLA should incorporate verification from previous versions, thus limiting the manual verification required for updating the final clinic list.

Two new primary care clinics in British Columbia are led by nurse practitioners (NPs) without FPs [49] and are, therefore, missing from the present clinic list. Additionally, rural and remote communities have clinics that may be serviced periodically by FPs, NPs, or registered nurses. It is unlikely that these locations were captured by the CPSBC Registry list. Should these types of clinics become more prevalent, the potential remedy for future versions of the final clinic list would be to apply a similar algorithm to registration lists of NPs and primary care registered nurses. Listing a hospital address was an exclusion criterion for the CLA. It is possible that in some settings, particularly rural areas, primary care is being provided at a hospital address. Future projects should verify if primary care is provided at the hospital address or if an FP registered with the CPSBC uses only a hospital address.

The utility of any registry of addresses relies on the currency and accuracy of the listing [48]. The CLA will require updating; fortunately, the physician registration list is updated continually by the provincial licensing body, and with appropriate resources, the algorithm can be repeated to update the accurate list of clinics.

Finally, the rapid growth of primary care services offered through virtual platforms introduces an important new element to this work [42,50]. FPs can now provide care to patients in distant locations. In British Columbia, virtual care via the single-payer system requires a physician to be registered with the CPSBC [51]. At this time, it is unclear how virtual services might impact the CLA.

Conclusions

The CLA can reliably create a list of primary care clinics based on publicly available information. The algorithm can be applied in other regions that need comprehensive lists of primary care clinics. The CLA offers researchers, decision-makers, and other organizations interested in health services a reliable way to estimate the regional distribution of primary care clinics. Future research could include the application of the CLA to the evaluation of initiatives for primary care transformation.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed description of the application of the Clinic List Algorithm (CLA) method.

[PDF File (Adobe PDF File), 240 KB - formative_v6i6e34141_app1.pdf]

Multimedia Appendix 2
Regular expression statements used in the identification and sorting of addresses from the College of Physicians and Surgeons of British Columbia list to each of the groups.
[PDF File (Adobe PDF File), 140 KB - formative_v6i6e34141_app2.pdf]

Multimedia Appendix 3
Accuracy of algorithm sorting of individual physician addresses from licence registration to identify physical locations of primary care clinics.
[PDF File (Adobe PDF File), 164 KB - formative_v6i6e34141_app3.pdf]

Multimedia Appendix 4
Internet validation process and phone call script.
[PDF File (Adobe PDF File), 108 KB - formative_v6i6e34141_app4.pdf]

Multimedia Appendix 5
Clinic list.
[PDF File (Adobe PDF File), 361 KB - formative_v6i6e34141_app5.pdf]

References

1. World Health Organization. Declaration of Astana: Global Conference on Primary Health Care. Geneva, Switzerland; 2018. URL: https://www.who.int/docs/default-source/primary-health/declaration/gcphc-declaration.pdf [accessed 2022-05-26]
2. Basu S, Berkowitz SA, Phillips RL, Bitton A, Landon BE, Phillips RS. Association of primary care physician supply with population mortality in the United States, 2005-2015. JAMA Intern Med 2019 Apr 01;179(4):506-514 [FREE Full text] [doi: 10.1001/jamainternmed.2018.7624] [Medline: 30776056]
3. Xu X. Nearly 900,000 British Columbians don't have a family doctor, leaving walk-in clinics and ERs swamped. The Globe and Mail. 2022 Apr 29. URL: https://www.theglobeandmail.com/canada/british-columbia/article-nearly-900000-british-columbians-dont-have-a-family-doctor-leaving/ [accessed 2022-05-11]
4. Physicians in Canada. Canadian Institute for Health Information. URL: https://www.cihi.ca/en/physicians-in-canada? [accessed 2021-03-04]
5. Lavergne M, Peterson S, McKendry R, Sivananthan S, McGrail K. Full-service family practice in British Columbia: policy interventions and trends in practice, 1991-2010. Healthc Policy 2014 May;9(4):32-47 [FREE Full text] [Medline: 24973482]
6. Stewart M, Ryan B. Ecology of health care in Canada. Can Fam Physician 2015 May;61(5):449-453 [FREE Full text] [Medline: 25971762]
7. Hutchison B, Levesque J, Strumpf E, Coyle N. Primary health care in Canada: Systems in motion. Milbank Q 2011 Jun;89(2):256-288. [doi: 10.1111/j.1468-0009.2011.00628.x] [Medline: 21676023]
8. Kiran T, Green ME, DeWit Y, Khan S, Schultz S, Kopp A, et al. Association of physician payment model and team-based care with timely access in primary care: A population-based cross-sectional study. CMAJ Open 2020;8(2):E328-E337 [FREE Full text] [doi: 10.9778/cmajo.20190063] [Medline: 32381684]
9. Leslie M, Khayatzadeh-Mahani A, Birdsell J, Forest PG, Henderson R, Gray RP, et al. An implementation history of primary health care transformation: Alberta's primary care networks and the people, time and culture of change. BMC Fam Pract 2020 Dec 05;21(1):258 [FREE Full text] [doi: 10.1186/s12875-020-01330-7] [Medline: 33278880]
10. Implementation of the Integrated System of Primary and Community Care: Team-Based Care Through Primary Care Networks Guidance to Collaborative Services Committees. Vancouver, BC: General Practice Services Committee; 2017 Dec. URL: https://gpscbc.ca/sites/default/files/PMH%20PCN%20CSC%20Guidance%20201712.pdf [accessed 2022-05-31]
11. Ministry of Health. Government of British Columbia. URL: https://www2.gov.bc.ca/gov/content/governments/organizational-structure/ministries-organizations/ministries/health [accessed 2022-05-30]
12. Somé NH, Devlin R, Mehta N, Zaric G, Sarma S. Team-based primary care practice and physician's services: Evidence from family health teams in Ontario, Canada. Soc Sci Med 2020 Nov;264:113310. [doi: 10.1016/j.socscimed.2020.113310] [Medline: 32877846]
13. Reiss-Brennan B, Brunisholz KD, Dredge C, Briot P, Grazier K, Wilcox A, et al. Association of integrated team-based care with health care quality, utilization, and cost. JAMA 2016;316(8):826-834. [doi: 10.1001/jama.2016.11232] [Medline: 27552616]
14. Meyers DJ, Chien AT, Nguyen KH, Li Z, Singer SJ, Rosenthal MB. Association of team-based primary care with health care utilization and costs among chronically ill patients. JAMA Intern Med 2019 Jan 01;179(1):54-61 [FREE Full text] [doi: 10.1001/jamainternmed.2018.5118] [Medline: 30476951]

15. Schoen C, Osborn R, Doty M, Bishop M, Peugh J, Murukutla N. Toward higher-performance health systems: Adults' health care experiences in seven countries, 2007. Health Aff (Millwood) 2007;26(6):w717-w734. [doi: 10.1377/hlthaff.26.6.w717] [Medline: 17978360]
16. Kash BA, Cheon O, Halzack NM, Miller TR. Measuring team effectiveness in the health care setting: An inventory of survey tools. Health Serv Insights 2018;11:1-18 [FREE Full text] [doi: 10.1177/1178632918796230] [Medline: 30158825]
17. Riordan F, McHugh SM, O'Donovan C, Mtshede MN, Kearney PM. The role of physician and practice characteristics in the quality of diabetes management in primary care: Systematic review and meta-analysis. J Gen Intern Med 2020 Jun;35(6):1836-1848 [FREE Full text] [doi: 10.1007/s11606-020-05676-7] [Medline: 32016700]
18. Oandasan IF, Archibald D, Authier L, Lawrence K, McEwen LA, Mackay MP, et al. Future practice of comprehensive care: Practice intentions of exiting family medicine residents in Canada. Can Fam Physician 2018 Jul;64(7):520-528 [FREE Full text] [Medline: 30002030]
19. Schultz S, Glazier R. Identification of physicians providing comprehensive primary care in Ontario: A retrospective analysis using linked administrative data. CMAJ Open 2017 Dec 19;5(4):E856-E863 [FREE Full text] [doi: 10.9778/cmajo.20170083] [Medline: 29259018]
20. Lavergne MR, Goldsmith LJ, Grudniewicz A, Rudoler D, Marshall EG, Ahuja M, et al. Practice patterns among early-career primary care (ECPC) physicians and workforce planning implications: Protocol for a mixed methods study. BMJ Open 2019 Sep 24;9(9):e030477 [FREE Full text] [doi: 10.1136/bmjopen-2019-030477] [Medline: 31551384]
21. Willis TA, West R, Rushforth B, Stokes T, Glidewell L, Carder P, ASPIRE programme team. Variations in achievement of evidence-based, high-impact quality indicators in general practice: An observational study. PLoS One 2017;12(7):e0177949 [FREE Full text] [doi: 10.1371/journal.pone.0177949] [Medline: 28704407]
22. Goldacre B, Reynolds C, Powell-Smith A, Walker A, Yates T, Croker R, et al. Do doctors in dispensing practices with a financial conflict of interest prescribe more expensive drugs? A cross-sectional analysis of English primary care prescribing data. BMJ Open 2019 Feb 05;9(2):e026886 [FREE Full text] [doi: 10.1136/bmjopen-2018-026886] [Medline: 30813120]
23. McLeod L, Buckley G, Sweetman A. Ontario primary care models: A descriptive study. CMAJ Open 2016;4(4):E679-E688 [FREE Full text] [doi: 10.9778/cmajo.20160069] [Medline: 28018882]
24. Wranik W, Price S, Haydt S, Edwards J, Hatfield K, Weir J, et al. Implications of interprofessional primary care team characteristics for health services and patient health outcomes: A systematic review with narrative synthesis. Health Policy 2019 Jun;123(6):550-563 [FREE Full text] [doi: 10.1016/j.healthpol.2019.03.015] [Medline: 30955711]
25. Shah TI, Bell S, Wilson K. Spatial accessibility to health care services: Identifying under-serviced neighbourhoods in Canadian urban areas. PLoS One 2016;11(12):e0168208 [FREE Full text] [doi: 10.1371/journal.pone.0168208] [Medline: 27997577]
26. Gilliland JA, Shah TI, Clark A, Sibbald S, Seabrook JA. A geospatial approach to understanding inequalities in accessibility to primary care among vulnerable populations. PLoS One 2019;14(1):e0210113 [FREE Full text] [doi: 10.1371/journal.pone.0210113] [Medline: 30615678]
27. Marshall EG, Gibson RJ, Lawson B, Burge F. Protocol for determining primary healthcare practice characteristics, models of practice and patient accessibility using an exploratory census survey with linkage to administrative data in Nova Scotia, Canada. BMJ Open 2017 Mar 16;7(3):e014631 [FREE Full text] [doi: 10.1136/bmjopen-2016-014631] [Medline: 28302637]
28. Partner organizations. Medical Council of Canada. URL: https://mcc.ca/about/partner-organizations/ [accessed 2021-04-05]
29. Registrant directory. College of Physicians and Surgeons of British Columbia. URL: https://www.cpsbc.ca/registrant_search [accessed 2022-05-26]
30. Community Health Service Areas - CHSA. Government of British Columbia Data Catalogue. Victoria, BC: Health Sector Information, Analysis and Reporting, British Columbia Ministry of Health; 2019. URL: https://catalogue.data.gov.bc.ca/dataset/community-health-service-areas-chsa [accessed 2022-05-26]
31. BC Address Geocoder. Government of British Columbia. URL: https://www2.gov.bc.ca/gov/content/data/geographic-data-services/location-services/geocoder [accessed 2022-05-26]
32. Walk-in Clinics in BC. Government of British Columbia Data Catalogue. Victoria, BC: HealthLinkBC, British Columbia Ministry of Health; 2022. URL: https://catalogue.data.gov.bc.ca/dataset/walk-in-clinics-in-bc [accessed 2022-05-26]
33. Howard M, Goertzen J, Kaczorowski J, Hutchison B, Morris K, Thabane L, et al. Emergency Department and Walk-in Clinic Use in Models of Primary Care Practice with Different After-Hours Accessibility in Ontario. Healthc Policy 2008 Aug;4(1):73-88 [FREE Full text] [Medline: 19377344]
34. Government of British. Hospitals in BC. Government of British Columbia Data Catalogue. Victoria, BC: HealthLinkBC, British Columbia Ministry of Health; 2022. URL: https://catalogue.data.gov.bc.ca/dataset/hospitals-in-bc [accessed 2022-05-26]
35. Government of British Columbia. Urgent and Primary Care Centres. Government of British Columbia Data Catalogue. Victoria, BC: HealthLinkBC, British Columbia Ministry of Health; 2022. URL: https://catalogue.data.gov.bc.ca/dataset/urgent-and-primary-care-centres [accessed 2022-05-26]
36. Long Term Care Directory – Information and Reports. Victoria, BC: British Columbia Office of the Seniors Advocate URL: https://www.seniorsadvocatebc.ca/long-term-care-directory/ [accessed 2022-05-25]

37. Innovation Support Unit. Clinic list BC. GitHub. 2021. URL: https://github.com/ISU-UBC/clinic-list-BC [accessed 2022-05-06]
38. Hedden L, Banihosseini S, Strydom N, McCracken R. Modern work patterns of "classic" versus millennial family doctors and their effect on workforce planning for community-based primary care: A cross-sectional survey. Hum Resour Health 2020 Sep 21;18(1):67 [FREE Full text] [doi: 10.1186/s12960-020-00508-5] [Medline: 32958028]
39. Reeve J, Beaulieu M, Freeman T, Green LA, Lucassen P, Martin C, et al. Revitalizing generalist practice: The Montreal Statement. Ann Fam Med 2018 Jul 09;16(4):371-373 [FREE Full text] [doi: 10.1370/afm.2280]
40. Kiran T, Victor J, Kopp A, Shah B, Glazier R. The relationship between primary care models and processes of diabetes care in Ontario. Can J Diabetes 2014 Jun;38(3):172-178 [FREE Full text] [doi: 10.1016/j.jcjd.2014.01.015] [Medline: 24909088]
41. McGrail K, Lavergne R, Lewis SJ, Peterson SLM, Barer M, Garrison SR. Classifying physician practice style: A new approach using administrative data in British Columbia. Med Care 2015 Mar;53(3):276-282. [doi: 10.1097/MLR.0000000000000310] [Medline: 25634088]
42. Matthewman S, Spencer S, Lavergne MR, McCracken RK, Hedden L. An environmental scan of virtual "walk-in" clinics in Canada: Comparative study. J Med Internet Res 2021 Jun 11;23(6):e27259 [FREE Full text] [doi: 10.2196/27259] [Medline: 34114963]
43. Dahrouge S, Hogg W, Younger J, Muggah E, Russell G, Glazier R. Primary care physician panel size and quality of care: A population-based study in Ontario, Canada. Ann Fam Med 2016;14(1):26-33 [FREE Full text] [doi: 10.1370/afm.1864] [Medline: 26755780]
44. Hedden L, Lavergne M, McGrail K, Law M, Bourgeault I, McCracken R, et al. Trends in providing out-of-office, urgent after-hours, and on-call care in British Columbia. Ann Fam Med 2019 Mar;17(2):116-124 [FREE Full text] [doi: 10.1370/afm.2366] [Medline: 30858254]
45. Desveaux L, Saragosa M, Kithulegoda N, Ivers NM. Family physician perceptions of their role in managing the opioid crisis. Ann Fam Med 2019 Jul;17(4):345-351 [FREE Full text] [doi: 10.1370/afm.2413] [Medline: 31285212]
46. Paré G, Raymond L, Castonguay A, Grenier Ouimet A, Trudel M. Assimilation of medical appointment scheduling systems and their impact on the accessibility of primary care: Mixed methods study. JMIR Med Inform 2021 Nov 16;9(11):e30485 [FREE Full text] [doi: 10.2196/30485] [Medline: 34783670]
47. Langton J, Wong S, Johnston S, Abelson J, Ammi M, Burge F, et al. Primary care performance measurement and reporting at a regional level: Could a matrix approach provide actionable information for policy makers and clinicians? Healthc Policy 2016 Nov;12(2):33-51 [FREE Full text] [Medline: 28032823]
48. Strumpf E, Levesque J, Coyle N, Hutchison B, Barnes M, Wedel R. Innovative and diverse strategies toward primary health care reform: Lessons learned from the Canadian experience. J Am Board Fam Med 2012 Mar;25 Suppl 1:S27-S33 [FREE Full text] [doi: 10.3122/jabfm.2012.02.110215] [Medline: 22403248]
49. BC Medical Journal. New nurse-practitioner primary care clinics. BC Med J 2020 Sep;62(7):250-251 [FREE Full text]
50. Glazier RH, Green ME, Wu FC, Frymire E, Kopp A, Kiran T. Shifts in office and virtual primary care during the early COVID-19 pandemic in Ontario, Canada. CMAJ 2021 Feb 08;193(6):E200-E210. [doi: 10.1503/cmaj.202303] [Medline: 33558406]
51. Practice Standard: Virtual Care. Version 8.3. Vancouver, BC: College of Physicians and Surgeons of British Columbia; 2022 May 06. URL: https://www.cpsbc.ca/files/pdf/PSG-Telemedicine.pdf [accessed 2022-05-24]

Abbreviations

CHSA: Community Health Service Area
CLA: Clinic List Algorithm
CPSBC: College of Physicians and Surgeons of British Columbia
FP: family physician
NP: nurse practitioner
UPCC: Urgent and Primary Care Centre

Edited by A Mavragani; submitted 21.10.21; peer-reviewed by D Madhusudhan, Y Xie; comments to author 18.03.22; revised version received 13.05.22; accepted 19.05.22; published 22.06.22.

Please cite as:
Cooper IR, Lindsay C, Fraser K, Hill TT, Siu A, Fletcher S, Klimas J, Hamilton MA, Frazer AD, Humphrys E, Koepke K, Hedden L, Price M, McCracken RK
Finding Primary Care—Repurposing Physician Registration Data to Generate a Regionally Accurate List of Primary Care Clinics: Development and Validation of an Open-Source Algorithm
JMIR Form Res 2022;6(6):e34141
URL: https://formative.jmir.org/2022/6/e34141
doi:10.2196/34141
PMID:35731556

Original Paper

Digital Health Apps in the Context of Dementia: Questionnaire Study to Assess the Likelihood of Use Among Physicians

Markus Schinle[1], MSc; Christina Erler[1], MSc; Mayumi Kaliciak[2], BSc; Christopher Milde[2], PhD; Simon Stock[3], PhD; Marius Gerdes[3], MSc; Wilhelm Stork[3], Prof Dr

[1]Medical Information Technology, Embedded Systems and Sensors Engineering, FZI Research Center for Information Technology, Karlsruhe, Germany

[2]Biological Psychology, Department of Psychology, University of Koblenz and Landau, Landau, Germany

[3]Institut fuer Technik der Informationsverarbeitung, Department of Electrical Engineering & Information Technology, KIT Karlsruhe Institute of Technology, Karlsruhe, Germany

Corresponding Author:
Markus Schinle, MSc
Medical Information Technology
Embedded Systems and Sensors Engineering
FZI Research Center for Information Technology
Haid-und-Neu-Str. 10-14
Karlsruhe, 76131
Germany
Phone: 49 721 9654 75
Email: schinle@fzi.de

Abstract

Background: Age-related diseases such as dementia are playing an increasingly important role in global population development. Thus, prevention, diagnostics, and interventions require more accessibility, which can be realized through digital health apps. With the *app on prescription*, Germany made history by being the first country worldwide to offer physicians the possibility to prescribe and reimburse digital health apps as of the end of the year 2020.

Objective: Considering the lack of knowledge about correlations with the likelihood of use among physicians, this study aimed to address the question of what makes the use of a digital health app by physicians more likely.

Methods: We developed and validated a novel measurement tool—the Digital Health Compliance Questionnaire (DHCQ)—in an interdisciplinary collaboration of experts to assess the role of proposed factors in the likelihood of using a health app. Therefore, a web-based survey was conducted to evaluate the likelihood of using a digital app called DemPredict to screen for Alzheimer dementia. Within this survey, 5 latent dimensions (acceptance, attitude toward technology, technology experience, payment for time of use, and effort of collection), the dependent variable *likelihood of use*, and answers to exploratory questions were recorded and tested within directed correlations. Following a non–probability-sampling strategy, the study was completed by 331 physicians from Germany in the German language, of whom 301 (90.9%) fulfilled the study criteria (eg, being in regular contact with patients with dementia). These data were analyzed using a range of statistical methods to validate the dimensions of the DHCQ.

Results: The DHCQ revealed good test theoretical measures—it showed excellent fit indexes (Tucker-Lewis index=0.98; comparative fit index=0.982; standardized root mean square residual=0.073; root mean square error of approximation=0.037), good internal consistency (Cronbach α=.83), and signs of moderate to large correlations between the DHCQ dimensions and the dependent variable. The correlations between the variables *acceptance*, *attitude toward technology*, *technology experience*, and *payment for the time of use* and the dependent variable *likelihood of use* ranged from 0.29 to 0.79, and the correlation between *effort of the collection* and *likelihood of use* was −0.80. In addition, we found high levels of skepticism regarding data protection, and the age of the participants was found to be negatively related to their technical experience and attitude toward technology.

Conclusions: In the context of the results, increased communication between the medical and technology sectors and significantly more awareness raising are recommended to make the use of digital health apps more attractive to physicians as they can be adjusted to their everyday needs. Further research could explore the connection between areas such as adherence on the patient side and its impact on the likelihood of use by physicians.

(JMIR Form Res 2022;6(6):e35961) doi:10.2196/35961

KEYWORDS

digital health applications; likelihood of use; usability; adherence; dementia; screening; treatment; physician; eHealth; questionnaire; mobile phone

Introduction

Background

The populations in Germany and the United States are characterized by a growing proportion of people aged >60 years [1,2]. As an age-related disease, Alzheimer dementia is of particular importance in our society [3,4] as approximately 5% of people aged >65 years already experience severe dementia, and another 10% experience mild to moderate dementia [5]. Owing to this demographic development, the number of patients with dementia is increasing [6,7]. As early detection of dementia offers the patient the opportunity to manage daily issues such as finances and insurance by themselves [8], it has led to the rising importance of dementia detection, especially in the early stages [9,10].

Dementia is a disease with various causes that entails an above-average loss of intellectual skills. The main group of people affected is in an advanced age of >65 years [5].

The course of dementia is usually chronic or progressive, affecting the higher cortical functions (eg, memory, orientation, cognition, learning, language, and judgment), motivation, social behavior, and emotional control of the person with the disease [11].

Approximately 60% of patients with dementia experience the so-called Alzheimer disease, which usually begins insidiously and leads to death after approximately 5 to 10 years [5].

To support the early detection of dementia, it is possible to use mobile solutions such as apps on smartphones or tablets. In this way, the financial burden on the health system can be reduced [12,13] as there is a time saving of up to 30% in comparison with traditional solutions [14]. There are further studies worldwide that also report time savings by using digital solutions (eg, in Austria [15], Germany [14,16-18], and the United States [12,13,19]), including the special case of dementia screening [16].

Owing to the ease of accessibility [14], many physicians and patients like to use smartphones for remote diagnosis [20]. Furthermore, physicians describe apps for diagnosis as the most useful implementation in the eHealth sector [21].

In addition, an interdisciplinary exchange is facilitated in the use of technological tools [22], which is good for research and provides a more accurate picture of diseases such as dementia because of larger data sets available and the possibility of monitoring patients longitudinally with a high temporal resolution [14]. Therefore, universities and the health care systems of every country should have a great interest in developing and promoting the use of technology [23].

Hereby, an improvement in treatment can be established [24,25], and medical apps can be considered very useful tools in prevention and therapy [15] as they are considered not only for use in diagnosis but also for remote monitoring of patients [21]. An example is monitoring patients with Alzheimer disease using the Android app iWander [4] to analyze their location using GPS to make life easier and minimize financial burden [20].

Apps such as these make it easier to care for patients with chronic diseases [21], and diagnosis apps permit early intervention, for example, through cognitive training for Alzheimer disease, which has been shown to have a positive impact on disease progression [26].

Prior Work

In Germany, approximately 1 in 5 physicians stated that they use technical aids in their contact with patients (eg, to support patient information or diagnostics), making this a frequently used device [27].

To analyze what exactly influences the use of technical aids, there are some studies in the United States in which it was found that a positive attitude toward smartphones or toward one's own ability to use them has a positive influence on the frequency of use [28,29].

Other groups of researchers have found that even observability, compatibility, job relevance, personal experience, and the internal and external environment influenced the attitude toward using a smartphone [30] and that the perceived usefulness also has an impact [29].

There are 3 groups of physicians: one-third are neutral, one-third welcome technical progress, and the other third is skeptical [31], especially concerning data security [18,22,24,25,27] and confidentiality [32,33], legal ambiguities, or a risk of abuse [18].

However, according to another study, there seem to be sophisticated approaches that increase transparency in apps for the user (eg, the physician) [13]. There are also strict rules for apps in Germany to be covered by health insurance via prescription as this requires a certification as a medical device [34] and also that safety can be ensured [35]. In the United States, it also depends on what the app is to be used for and which rules apply to it. For example, medical apps need approval from the Food and Drug Administration or a certification from private sources; however, these certifications are expensive [13].

Therefore, many app developers circumvent these often long-lasting certification processes by not considering the billing of health insurance as an option or by declaring their app only as an accompaniment but not a diagnostic instrument. The latter explains why there are some apps on the market for digital dementia screening that often do not fulfill medical criteria, such as those provided by the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition [36], for testing complex attention, executive functions, learning, and memory or language [37,38], which explains the physicians' demand for a binding test seal [39].

Many app developers also disclose little information in the apps [40]. An attempt to establish a reliable seal for consumers in Germany is the *HealthOn-Apps Ehrenkodex* (code of ethics), which is intended to provide guidance by checking the app for a data protection notice, author information, source information, freedom from advertising, financing information, contact data, and imprint [40].

Another point to consider when using apps in medicine is the age of the physician and the patient. Older patients usually make more frequent visits to physicians because of age-related conditions but are the least likely group to use smartphones [13]. For example, in Austria, approximately 64% of younger individuals (aged <44 years) and only approximately 39% of older individuals (aged >44 years) use a smartphone [15]. Unfortunately, however, age and age-related limitations are often not considered by developers, which leads to avoidable difficulties in use [27,41].

This is happening even though it has long been known that differences in attitudes toward technology can be attributed to the age and educational level of the participants (physicians and patients) [27]. For example, a Swiss study identified a group of younger participants among physicians with a positive attitude [42], and a German study found that younger participants rated the opportunities for digitization significantly higher (aged 20-29 years: 93% compared with aged >70 years: 44%) [18].

Goal of This Work

There is limited research worldwide on the acceptance of mobile health technologies among physicians [43], and the mass of offerings (apps) is difficult to sort through [44]. In addition, most research is limited to the patient side [45], which is why more attention should be paid to the needs of physicians in the development that is already taking place [13,34]. To this end, a collaboration between behavioral researchers and developers is of great importance and could lead to better results concerning the health care system [13].

When it comes to the question of how exactly an app must be designed so that physicians are willing and motivated, the factors that correlate with the likelihood of use should be examined more closely. A well-designed app could support the diagnosis, disease monitoring, and treatment of patients with chronic diseases, especially in low-income countries [13]. Moreover, medicine in general could achieve more efficient information exchange and collaboration [14,22] and, thus, better care [24,25].

The test theoretical evaluation of the Digital Health Compliance Questionnaire (DHCQ) using confirmatory factor analysis (CFA) is to be defined as an initial goal whereby the latent factors, as well as the items that make up these factors, were selected by an interdisciplinary expert committee. A further objective was to answer the guiding research questions of (1) which *physician needs* are related to a high likelihood of using health apps and (2) which *physician characteristics* (skills and attitude toward technology) are correlated with an increased likelihood of using health apps. In addition, it is an example of interdisciplinary collaboration, thus uncovering skepticism and using it as a basis for better education to be able to accomplish (as recommended [10]) more early Alzheimer diagnoses and, thus, increase the rate of affected individuals who are treated (currently at only approximately 50% [14]).

Hypotheses

It is to be examined which dimensions, in general, are related to physicians' likelihood of use. On the basis of the aforementioned indications from the literature and an expert panel consisting of psychologists, geriatricians, pedagogues, and engineers with years of experience in the field of eHealth, the factors to be investigated were defined. For this purpose, *attitude toward technology*, *technology experience*, *payment for the time of use*, and the *effort of the collection* will be considered as independent variables, and *likelihood of use* will be considered as a dependent variable. To be able to apply the data to an example, special items were developed that capture the independent variable *acceptance of the app*. For this purpose, the functionality of the digital dementia screening app DemPredict [16,46] was briefly explained in a free field of this study, and some questions on its acceptance were formulated.

From these 6 dimensions, five hypotheses can be derived: (1) the likelihood of using the app is positively related to its *acceptance* (hypothesis 1), (2) the likelihood of using the app is positively related to the *attitude toward technology* (hypothesis 2), (3) the likelihood of using the app is positively related to *technology experience* (hypothesis 3), (4) the likelihood of using the app is positively related to the *payment* for the time of use (hypothesis 4), and (5) the likelihood of using a survey method is negatively related to the *effort of collection* (hypothesis 5).

In addition to the main hypotheses, secondary research questions were posed, which will be discussed in more detail in the Results and Discussion sections: (1) Does *age of the test person* correlate with *technology experience* as well as *attitude toward technology*? (2) Does it matter whether the app is declared as a medical device? (3) What role do concerns about data protection play? (4) Do physicians think that the time required for a dementia screening via app is higher for older people? (5) Is digital support for early monitoring of disease progression desired? (6) Can physicians imagine having a screening carried out under the supervision of a physician's assistant or would they even trust a result brought from home or a test carried out alone in the waiting room?

Methods

Research Design

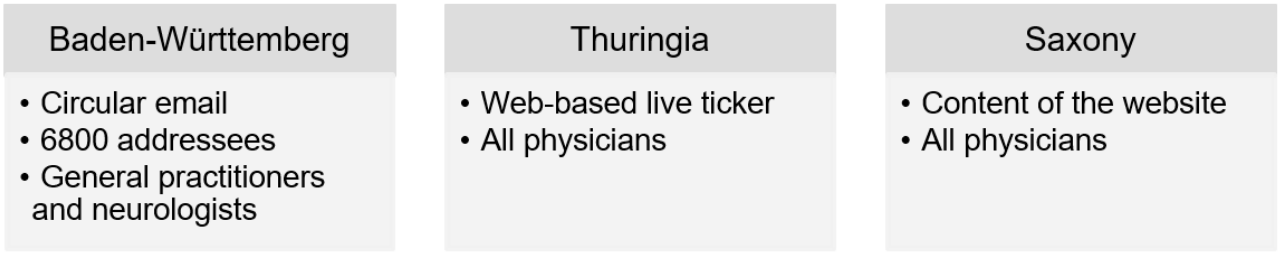
The data collection was implemented as a web-based survey using the tool SoSci Survey (version 3.2.21; SoSci Survey GmbH) and was conducted in February and March 2021. After adjustment, 301 participants remained to be included in the evaluation (*Sampling Design and Recruitment*). Statistical analysis was conducted using the statistical programming language R (version 4.0.4; R Foundation for Statistical Computing) as implemented in the R Studio environment. The CFA was used with the packages lavaan [47] and semPlot [48] using the commands *sem* and *semPaths*. Correlational hypotheses were tested using the command *cor.test* and illustrated using the package ggplot2 [49].

Sampling Design and Recruitment

The required sample size was estimated using G*Power (version 3.1.9.7; University of Düsseldorf) assuming an α of .05 and a power of 1 – β of .80. Despite intensive research, we could not find any comparable studies that would have made a priori sample size planning possible. The effect size was estimated by the consortium to be small (r=0.1 according to Cohen [50]) to obtain a feeling for the implementability of the study. This resulted in a sample size of 67 participants. Compensating for an attrition rate of 15%, we had to acquire ≥78 participants.

To recruit physicians, we followed a non–probability-sampling strategy and contacted all of the 16 regional Associations of Statutory Health Insurance Physicians (Kassenärztliche Vereinigungen) in Germany, the Association of General Practitioners (Hausärzteverband), the German Medical Association (Bundesärztekammer), and the German Society for General Practice and Family Medicine e.V. On the basis of these initial contact requests, a successful contact was established with the Associations of Statutory Health Insurance Physicians of the states of Baden-Württemberg, Thuringia, and Saxony. The physicians were recruited through the channels shown in Figure 1.

Figure 1. Recruiting channels.

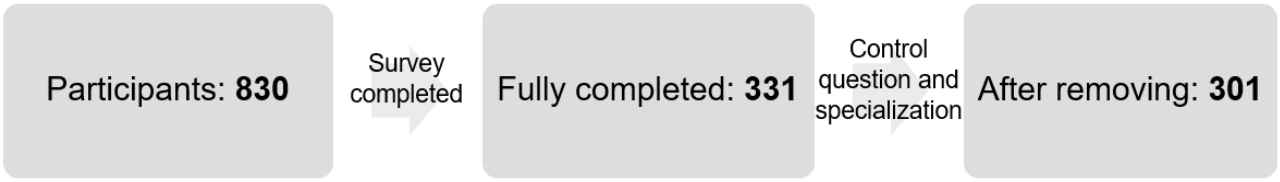


Study Sample

A total of 830 physicians participated in the study (Figure 2). Physicians who did not complete the survey were excluded from the study (fully completed: 331/830, 39.9% of the participants). In addition, there were 2.3% (19/830) of participants who only insufficiently or not at all completed the control question *To ensure the evaluability of the study, please briefly describe in your own words what the DemPredict application is supposed to do* (test participants 55, 147, 218, 224, 279, 360, 369, 417, 419, 452, 468, 480, 541, 593, 601, 647, 656, 709, and 731).

Furthermore, 11 physicians from nonrelevant specialties (definitely not in contact with patients with dementia or involved in dementia diagnosis) worked on the study (test participant 61, research associate; test participant 330, cardiology; test participant 574, orthopedics; test participant 587, dermatology; test participant 641, surgery; test participant 649, oncology; test participant 771, pediatrics; test participant 786, urology; test participant 797, gynecology; test participant 837, trauma surgery or orthopedics; and test participant 870, anesthesiology). They were also subsequently removed from the data set.

Figure 2. Study sample.



Ethics Approval

The study was approved with ethical approval number LEK-319 (application date April 02, 2021) by the local ethics committee of the University of Koblenz-Landau, Germany. The study is replicable because of the transparent design of the questionnaire. All participants were informed of the voluntary nature of their participation, their anonymity, and the possibility of asking questions. They agreed to the privacy policy by marking a checkbox in the web-based questionnaire.

Data Collection: Factors Determining Likelihood of App Use

Data collection was performed with a designed questionnaire called the DHCQ (Multimedia Appendix 1 [26]), which was partially based on items from other studies [51-53]. It is divided into 7 sections. In addition to demographic data, the survey asked about technology experience and attitudes toward technology. The questionnaire encompassed open- and close-ended or multiple-choice items. Information on the app DemPredict was followed by questions about attitudes toward this app and the requirements that such an app must fulfill to be used successfully in standard care.

The 2 dimensions *attitude toward technology* and *technology experience* are common dimensions in the literature [51-53], to which only new items were added. To define the new dimensions *acceptance*, *survey effort*, and *payment*, an interdisciplinary expert committee of psychologists, geriatricians, pedagogues, and engineers was formed, which discussed the items for these new dimensions.

To verify the proper understanding of the survey items and to reduce additional errors such as spelling or formatting errors, several pilot data sets were acquired within the research group by 3 members of the medical staff and 3 engineers beforehand.

Statistical Analysis

Overview

The 5 main hypotheses were all directly correlational. Thus, 1-sided Pearson correlations with a significance threshold of *P*=.05 were used to assess statistical significance. The Pearson correlation coefficient (*r*) was used as an effect size measure.

Item Analysis

Descriptive statistics were used to assess the suitability of the items and scales for the subsequent CFA and correlational analysis. Therefore, we computed measures of location (arithmetic mean and median), dispersion (SD), and shape (skewness and kurtosis; Tables S1 and S2 in Multimedia Appendix 2). To assess the normality of the items, a skew value of >2 or kurtosis of >7 was considered nonnormal [54]. The level of discrimination should be >0.3 to be *good* and >0.39 to be *excellent* [55]. Difficulty in this questionnaire was not similar to measuring the difficulty of solving an item but can be defined as *endorseability* [56] because a Likert scale was used. A difficulty range of 30 to 60 was *good*, a lower range was *difficult*, and a higher range was *easy* [57].

To interpret the *P* values from the Pearson correlations for the main hypothesis, the assumption of bivariate normality must be fulfilled. To test this assumption, the Shapiro-Wilk test was used. We used bias-corrected and accelerated bootstrapped Pearson correlations to obtain robust correlation statistics (1000 iterations).

CFA Component

To test the validity of the measurement model and to illustrate correlations, we decided to use a CFA to support the previous theoretical considerations.

The graphical representation of the parameters is shown in Figure 3.

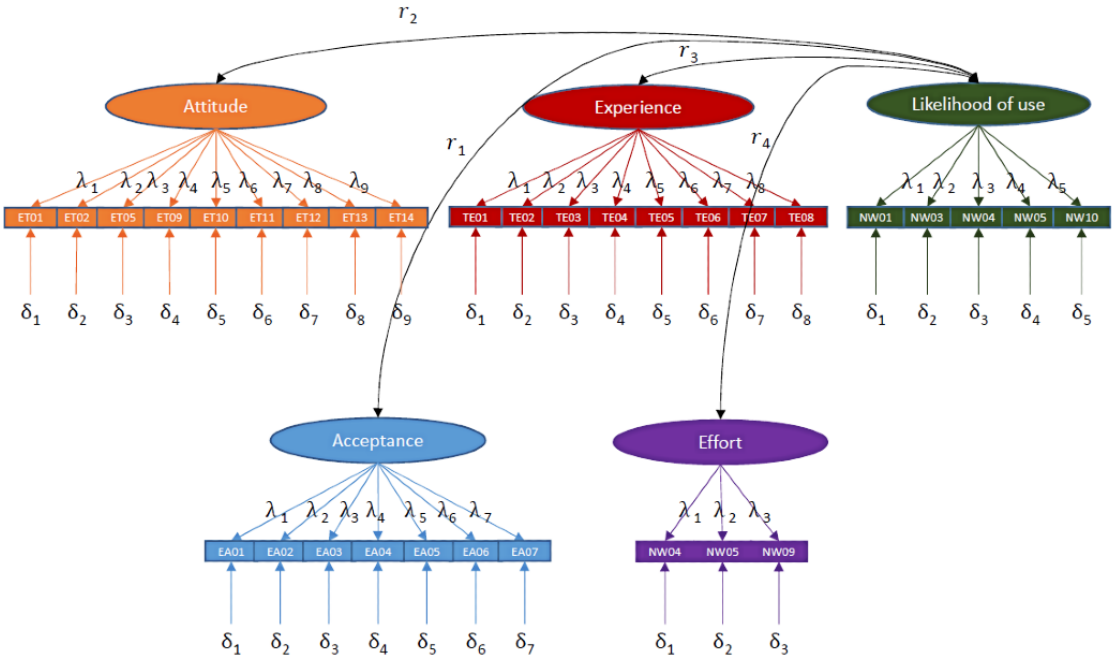
From this, we can define formulas, of which one is shown as follows using 1 item as an example: $EA01 = \lambda 1 \times acceptance + 0 \times attitude + 0 \times experience + 0 \times effort + 0 \times likelihood\ of\ use + \delta 1$.

The CFA is overidentified as, with 68 unknown variables—loads (λ; 32) + error variances (δ; 32) + correlations (*r*; 4)—a total of 528 known variables can be identified (with n=32). Thus, parameter estimation and goodness-of-fit testing were possible. To check whether the model-theoretical covariance matrix differed significantly from the observed matrix, parameter estimation was performed [58]. Owing to the very coarse scale (values from 1 to 5), the skewness of the items and scales [59] (*Item and Scale Analysis*), and the ordinal scaling of the items [60], parameter estimation should not be performed using the maximum likelihood method for these data. Nevertheless, to be able to test the a priori assumption regarding the latent constructs (measurement model), we decided to perform parameter estimation using the diagonally weighted least squares (DWLS) estimator, which is more robust for nonnormally distributed and ordinal data [60], to be able to evaluate the model fit in the next step.

To describe the fit of a model, fit indexes such as the standardized root mean square residual, the root mean square error of approximation, the comparative fit index (CFI), and the Tucker-Lewis index (TLI) were used. For the badness-of-fit measures (standardized root mean square residual and root mean square error of approximation), lower values (approximating 0) indicated a high model performance and, for the goodness-of-fit measures (CFI and TLI), higher values (approximating 1) indicated a high model performance [61]. As the model chi-square is not very reliable for large samples, especially with skewed item distributions, it will not be considered as a model performance measure in our sample because the dependence of the probability of error β on the sample size [58] quickly results in a significant value for 301 participants.

Everything needed to determine validity is found in the output of the CFA [62,63]. Convergent validity can be described by good factor loadings, which do not cross-load on nonrelevant constructs [64]. Both convergent and discriminant validity are represented in good model fit indexes (in this study, CFI and TLI) [61]. Internal consistency was assessed using the Cronbach α [64], with α>.70 being recommended and α>.90 considered as redundant [64-66].

Figure 3. Measurement model. r: correlations; λ: loads; δ: error variances; ET: Attitude toward Technology; TE: Technical Experience; NW: Probability of Use; EA: Attitude toward App Acceptance.



Results

Overview

The results of this study are presented in the following sections. First, the distribution of the demographic data in the sample is shown, followed by a description of the item and scale scores and the results of the CFA. After that, we examine the reliability and validity of this study and, finally, explain the results of the hypotheses and exploration.

Sample Statistics

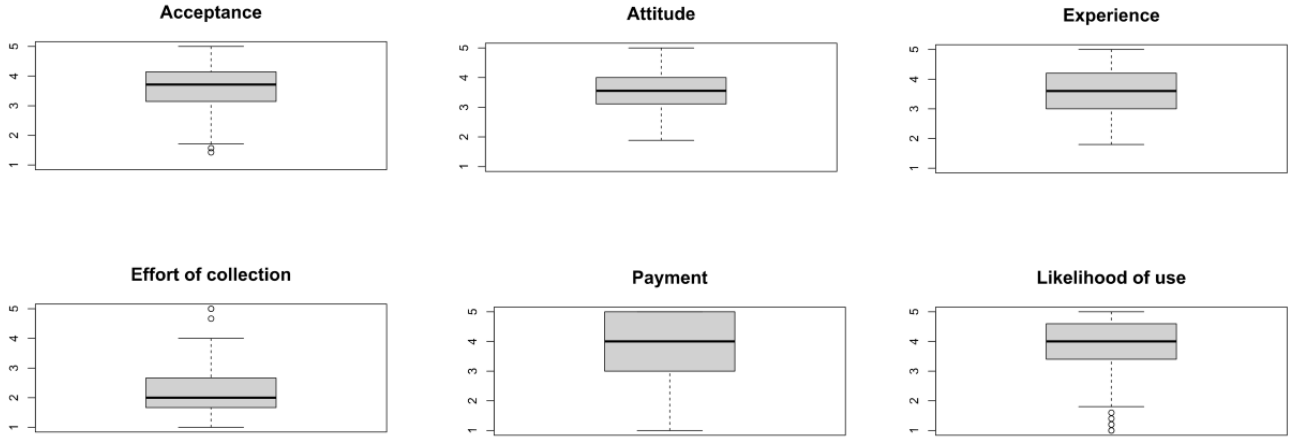
The age of the participants was approximately recorded by age groups to simplify the evaluation. This ranged over 4 groups specified in years (<30 years, 30-45 years, 45-60 years, and >60 years), which is why it is not useful to specify the mean and SD.

When specifying the gender, there were 3 selection options in the questionnaire (male, female, and diverse), and it can be said that the sample was balanced regarding gender types *Male* and *Female*. In this study, the specialty of the attending physician played a role as domain knowledge about dementia should be present, which is why it was included in the survey. All sociodemographic data are shown in Table S3 in Multimedia Appendix 2.

Item and Scale Analysis

All measures of central tendency and dispersion can be found in tables with item and scale values in Multimedia Appendix 2. It is noticeable that the items and scales partly show skewness. Moreover, as the data are ordinally scaled and the scale is only 5 levels, the methodology of the CFA was adapted (DWLS instead of maximum likelihood as parameter estimation). The difficulty range in 12 items was between 30 and 60 and, in the other items, between 60 and 82. Regarding the fact that items such as *Electronic devices make my everyday life easier* or *A wisely designed application can support an anamnesis just as well as a paper test* clearly are easier to respond to, these were acceptable values to continue the evaluation [56,57]. The level of discrimination in all items except the item that represents the effort for testing with tools such as DemPredict under the supervision of an assistant (item NW09.m: 0.08) was between 0.27 and 0.79, which are good to excellent values (Table S1 in Multimedia Appendix 2). In addition, the scales were tested for their normal distribution using the Shapiro-Wilk test to check the characteristic values. It can be seen that they are not normally distributed except for the dimension *attitude toward technology* (Table S2 in Multimedia Appendix 2). However, owing to the sample size of 301 participants, the data could be evaluated and interpreted [54,58]. As shown in Figure 4, the dimensions *acceptance*, *attitude*, *experience*, *effort of collection*, and *likelihood of use* were all checked for outlier values. No abnormalities were found either.

Figure 4. Box plots of dimensions from does not apply (1) to applies (5), inverted for effort.



CFA Component

Now that the measurement model has already been specified, the model fit is examined in more detail and the results are presented. After specifying the measurement model by graphical representation, setting up the equations, and checking the identifiability as well as the parameter estimation using the DWLS estimator, the overall model fit (Table 1) can be judged as *acceptable* [67].

Thus, the measurement instrument can be classified as functional and used as a basis for further data processing. The presentation of the results of the CFA is shown in Figure 5.

At this point, the correlations of the latent dimensions can be considered, which will be further explored in the follow-up by testing the hypotheses. It can be observed that all items show a satisfactory to very good loading and—matching the acceptable to very good fit indexes—correctly reflect the respective dimensions (Figure 5). A good construct convergent validity was indicated by the high factor loading, which can be classified as *good* (>0.55) [64,68]. Moreover, the goodness-of-fit indexes CFI and TLI (Table 1) indicated a *very good* overall model fit [61] and were also a sign of good discriminant validity [64,69]. As the instrument could generalize well [64] in specialties other than dementia, a good external validity was also expected. It can be used for every health app with only minor adjustments—such as in the dimension *acceptance of the app*, which was adapted to the app DemPredict (*Hypotheses* and Schinle et al [16,46]). The Cronbach α indicated a *very good* reliability with a value of .83 (95% CI 0.8-0.86) [65]. The power of the study in the total sample with an α error level of .05 and a sample size of 301 participants was 1. As there is a relationship between power, sample size, and postulated effect size and because of the dependence of the probability of the β error on sample size [58], the interpretability of the significance of the results was questioned, leading us to take the following further step: using 4 randomly generated subsamples, the power was recalculated. Now, a slightly different picture emerges from the overall sample. With a sample size of 75 and an α error level of .05, the average power was 0.966. As, in a Pearson correlation, the correlation coefficient *r* represents the effect size, the effect sizes are also reported and can be used by a replication study or a study in a similar field in a priori sample size planning.

Table 1. Model fit indexes.

Indexes	Actual value	Set point
RMSEA[a]	0.037	<0.08
SRMR[b]	0.073	<0.10
CFI[c]	0.982	>0.90
TLI[d]	0.980	>0.95

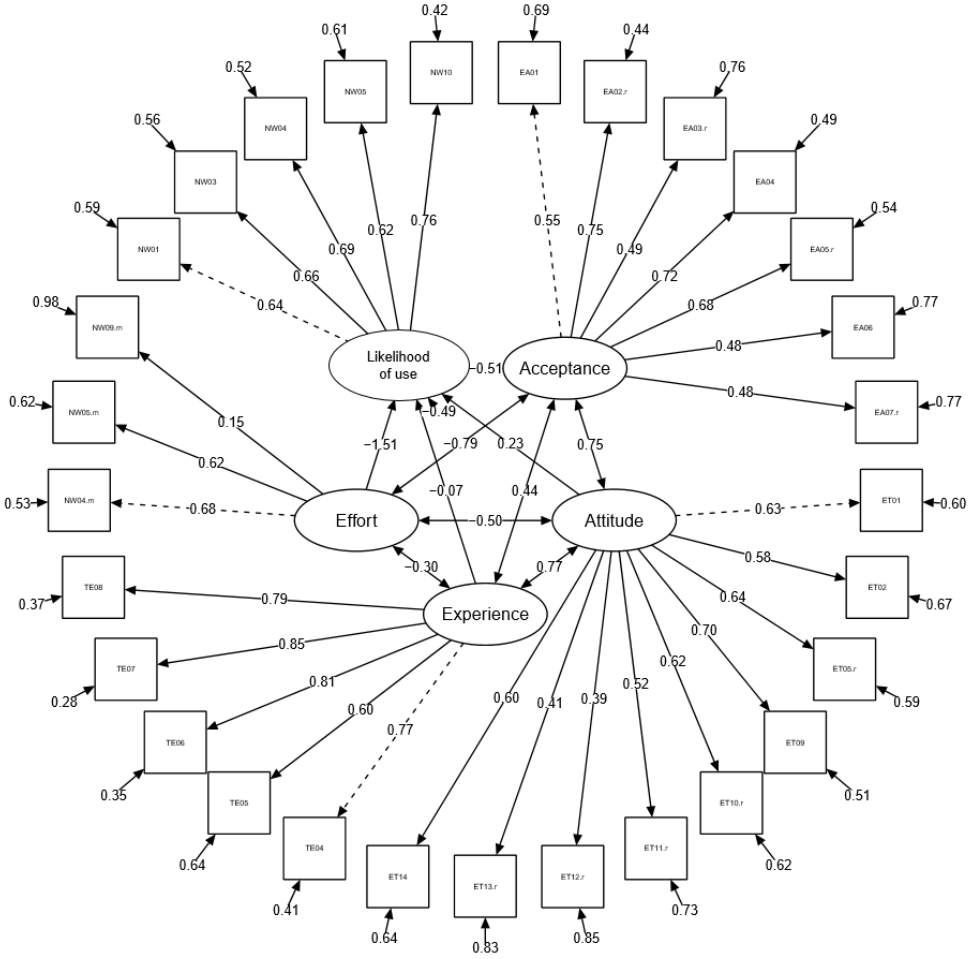
[a]RMSEA: root mean square error of approximation.

[b]SRMR: standardized root mean square residual.

[c]CFI: comparative fit index.

[d]TLI: Tucker-Lewis index.

Figure 5. Results of confirmatory factor analysis for structured equation model. ET: Attitude toward Technology; TE: Technical Experience; NW: Probability of Use; EA: Attitude toward App Acceptance.



Evaluation Outcomes

Main Hypotheses

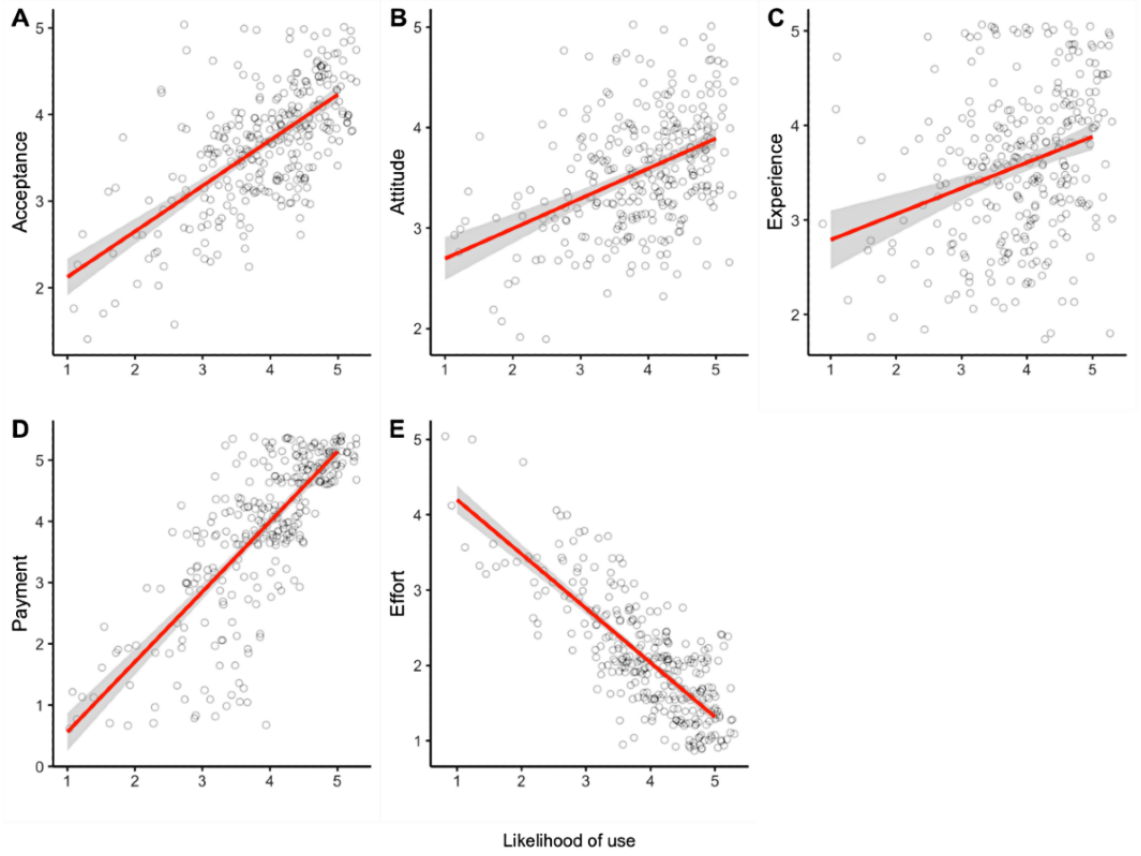
All hypotheses were tested for robustness (Table 2). For this purpose, previous analyses were performed on the normal distribution (*Item Analysis*). As the sample with n=301 is larger than n=30, it can be assumed that the values are nevertheless robust despite not being normally distributed, and significance tests can be performed [58].

It can be stated that all 5 hypotheses can be confirmed with good significance values (also considering the sample size) for this purpose compared with the 4 subsamples (see power). All correlations are in a medium- to high-value range and, thus, show clear correlations (Figure 6; Figures S1-S5 in Multimedia Appendix 2).

Table 2. Statistical values of the hypotheses.

Factors	1-tailed *t* test (*df*)	*P* value	Correlation (95% CI)	Effect
Acceptance+likelihood of use	14.822 (299)	<.001	0.65 (0.59 to 1.00)	Large
Attitude+likelihood of use	8.356 (299)	<.001	0.44 (0.35 to 1.00)	Medium-large
Experience+likelihood of use	5.228 (299)	<.001	0.29 (0.20 to 1.00)	Medium
Effort+likelihood of use	22.09 (299)	<.001	0.79 (0.75 to 1.00)	Large
Payment+likelihood of use	−22.97 (299)	<.001	−0.80 (−1.00 to –0.76)	Large

Figure 6. Correlations between items and likelihood of use (A-E).

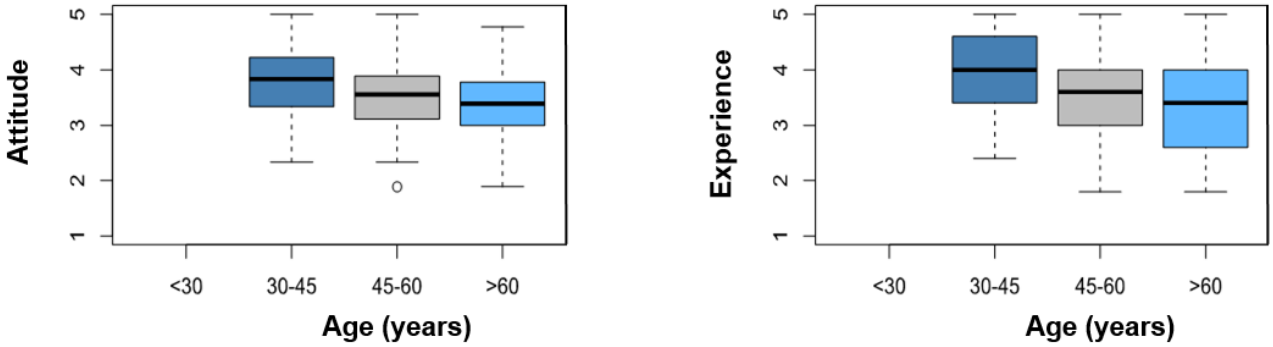


Secondary Research Questions

Correlation of the Age of the Test Person With Technology Experience as Well as Attitude Toward Technology

As Figure 7 clearly shows, there is a correlation between technical experience and age, and it can be observed that the median decreases with increasing age. As the graph of attitude toward technology is very similar, it can be assumed that there is a negative correlation between experience and attitude and age (ie, experience decreases and attitude deteriorates with increasing age).

Figure 7. Box plots of attitude and age and of experience and age.



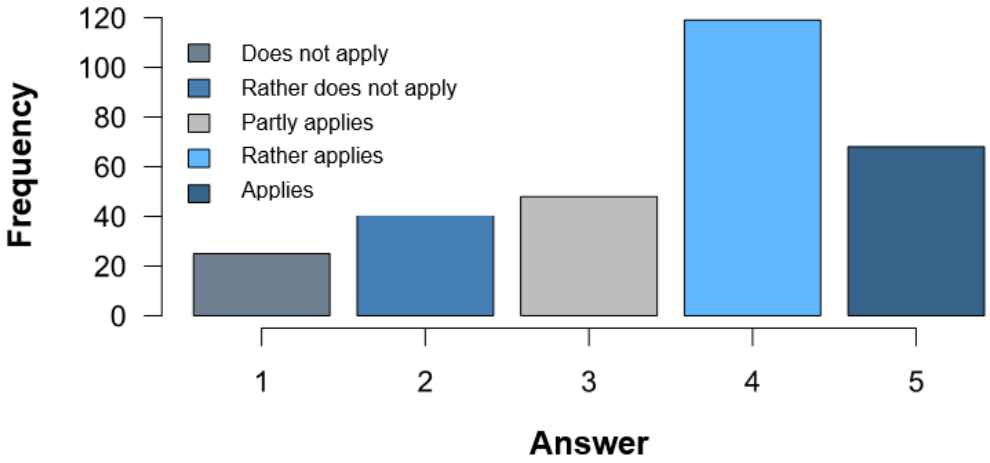
Effect of the App Being Declared as a Medical Device

The mean value for item NW07—*If the application is certified as a 'medical device,' this increases the likelihood of using it*—was 3.38 (SD 1.29) on a scale of 1 (does not apply) to 5 (applies), which is well above the mean. The modal value of 4 (rather applies) also indicates the clear direction that an app declared as a medical device is better accepted (Figure S6 in Multimedia Appendix 2).

Role of Concerns About Data Protection Play

Item EA03—*I have concerns about problems with data protection*—had a mean value of 3.54 (SD 1.21) on the same scale of 1 to 5, which is also well above the mean. Here, the modal value of 4 also shows a clear tendency toward worries about data protection. As data protection is a very important factor, Figure 8 underlines this statement.

Figure 8. Concerns about data protection.



Physicians' Thoughts About Whether the Time Required for a Test via App Is Higher for Older People

Item EA05—*I think it is more time-consuming to use an app with older people than a paper test*—showed a disagreement among physicians, which is also reflected in the mean value (2.97, SD 1.14). Answer option 3 (partly applies) received the most votes, and a symmetrical picture of the remaining answer options was formed whereby critical and uncritical votes balanced each other out (Figure S7 in Multimedia Appendix 2).

Need of Digital Support for Early Monitoring of Disease Progression

The question of whether they would like digitally supported monitoring of the course of the disease (item NW12) was answered by 86% (259/301) of the physicians with *yes*. Among the 301 votes, there were 4 (1.3%) abstentions (Figure S8 in Multimedia Appendix 2).

Physicians' Concerns about Having a Screening Carried Out Under the Supervision of a Physician's Assistant or the Result Brought From Home or a Test Carried Out Alone in the Waiting Room

A total of 48.5% (146/301) of the physicians could not imagine relying on a test result brought in by the patient. Approximately one-third (102/301, 33.9%) of the participants answered the question with *maybe* and partly filled in a free-text field provided with reasons. Processing in the waiting room without supervision also encountered skepticism. The mean value of the question of whether the physicians would allow the test to be performed in the waiting room without supervision was 2.64 (SD 1.22), and the most frequently selected answer was *rather does not apply* (Figure S9 in Multimedia Appendix 2). However, they were very open to the idea of having the test performed under the supervision of a physician's assistant to save time on the part of the physician. Most respondents decided to select the answer *rather applies* (122/301, 40.5%) or *applies* (59/301, 19.6%). The high mean score (3.52, SD 1.17) also indicates a clear direction in the responses (Figure S10 in Multimedia Appendix 2).

Discussion

Principal Findings

As more and more older people live in Germany [1] and the United States [2], age-related diseases such as dementia [3-5,7,9] will continue to increase. Thus, advancing technology to support early diagnosis and low-threshold access to care is becoming more important [8-10,22].

However, for a novel technology to be accepted and more likely to be used, it must be adapted to the different stakeholders. This led us to the guiding research question of what physicians need for the use of an app and what correlations exist on the part of skills and attitudes toward technology.

It turned out that the acceptance of this app was particularly important, with a correlation of 0.65. Attitude toward technology also played a decisive role (r=0.43), followed by technology experience (r=0.29). However, by far the most important factors were payment for the time of use (r=0.79) and the effort of the collection (r=−0.80).

Accordingly, the results show high correlations between the latent dimensions and the probability of use by the treatment providers. The expectations of the results were met, the theoretical considerations could be substantiated with partly very high correlations, and the significance was in a very good range even in smaller subsamples. This again allows for a very reliable interpretation and strengthens the importance of the investigated dimensions. From this, conclusions can be drawn for researchers and the need for action can be specified.

Comparison With Prior Work

To further increase the acceptance of individual apps, efforts need to be more targeted to meet the requirements from physicians' perspectives. The physicians would like the apps to be easy to use and evaluate and would also like the results to be presented in a comprehensible way (see the free-text fields in Multimedia Appendix 1). The free-to-use app of the World Health Organization—the mhGAP Intervention Guide for mental, neurological, and substance use disorders in nonspecialized health settings from the Mental Health Gap Action Programme—could be a low-threshold entry point [70].

However, there is still a lack of communication between developers and practitioners, which was partly reflected in frustration in this study. Answers in free-text fields, such as "More sales than science, though?" (test participant 227) or "Many people want to make money with physicians..." (test participant 204), show this great skepticism on the part of the medical profession (as well as *Commercialism and Wild West Relationships* [44]) but are counterbalanced by comments such as "I expressly welcome your efforts, as the coming generations will certainly be accustomed to digital formats" (test participant 375), which also fits with the year-on-year comparison in 2016 to 2017, in which the rates for "I think the development is good..." increased significantly from 12.5% to 23.5% [24].

Thus, despite skepticism, it can be assumed that there is an openness among treatment providers to engage with the technology of a specific app, which is evident in comments such as "I would apply it [...] if colleagues report positive experiences." This also fits with results from previous studies, where 42.6% of respondents voted for "I think the development is good in principle, but wait until there is more experience with it" [24], and there is a fundamentally positive attitude among physicians [42].

Furthermore, concerning the current skepticism, hope can be placed in a generational change among the treating physicians as our data showed a correlation between the age of the participants and their attitude toward technology or their experience with technology. This can be explained by the fact that younger participants have already grown up with technology, which also fits with a Swiss study ("Among the very positively attuned physicians, a group with a high proportion of younger and stationary physicians could be identified" [42]) and other studies and articles worldwide [13,15,18,27,41]. Therefore, technology experience may become less important in the future because of the postmaturing generation of the physician workforce.

To close this gap, consideration should be given on the part of policy makers but also by developers to make the transition to digital methods easier and more attractive. For example, in other free-text fields in our study, physicians suggested offering a 30-day trial or providing better information and making the switch easier using explanatory videos, webinars, information sheets, or demonstration versions.

Nevertheless, a well-adjusted app has great potential to be used even in older groups of physicians because of saved effort and, therefore, also better payment (hypotheses 4 and 5). The time-saving argument was an important factor in our study, as in many others [12-19].

The answers to the exploratory questions also provide a picture consistent with the literature. For most physicians, an app must be declared as a medical device. This supports other surveys in which a binding test seal is demanded [39] and is one of the prerequisites for billability with the health insurance fund in Germany [34] and approval by the Food and Drug Administration in the United States [13], which in turn is considered important by most practitioners for the use of an app. Developers should focus on certification as a medical device, for example, because of the additional safety [35] or because of the health care system requiring it [13,34].

The question of whether physicians have concerns about data protection was predominantly answered with *rather applies*, which matches the results of other studies [18,22,24,25,27,32,33], indicates a great need for more education regarding the technology and its safety, and supports efforts such as a seal for consumers [40].

Limitations

It becomes clear that developers of apps must pay attention to the relationship between the attitude toward technology and technology experience by the target group of physicians to make any age-appropriate adjustments that make it easier for physicians to use the app. Here, a less coarse division of the age groups would have been useful for a better assessment of the correlation between age and technology experience, which should be considered in further studies on this topic.

As general practitioners play an important role in dementia diagnoses because of their proximity to the patient [8], this study focused predominantly on them. However, these physicians are usually not specialists, and the results are not transferable to the entire medical profession. This has an impact on the final results as it can be assumed that some test persons are not as familiar with the clinical picture of dementia as specialists in the fields of neurology or geriatrics.

It would be beyond the limits of this work to survey physicians throughout Germany or even Europe or the world as a large proportion of participants were recruited via a circular mail from the regional Association of Statutory Health Insurance Physicians of the German federal state Baden-Württemberg. This results in a strong overrepresentation of participants from this federal state and, therefore, the results, assuming differences between east and west or north and south, cannot be generalized to Germany or beyond as a whole.

Outlook

Research should be significantly expanded in the field of dementia screening app development to further reduce existing skepticism [18,22,24,25,27,32,33] and increase confidence in the technology. It should be noted that misdiagnosis using poorly developed apps is dangerous [8] and can have far-reaching consequences.

Some physicians also worry about an increase in time spent using apps to diagnose dementia in older people (*Secondary Research Questions*). As the samples in other studies on the DemPredict project were far too small to make a statement about the behavior of older people when using the app, we recommend further research in this field to positively encourage the medical profession to also rely on the technical assistance of the app for dementia diagnoses in older patients provided that it has been tested on older and already affected patients [46,53].

Further research is needed based on the study result that digital support is desired for early monitoring of the course of the disease in patients with dementia and the uniform picture that physicians can imagine having a digital test performed by a physician's assistant (*Secondary Research Questions*). Here,

elaboration and further interviews are necessary to create a clearer offer for the use of these 2 options.

Sentences such as “I don’t know if my clientele will engage with an app. They always look for a personal conversation with the doctor” (test participant 115) indicate a problem not considered in this study: acceptance (adherence) on the patient side, which of course also influences the use of the app by the physician. Here, future research should not lose sight of the connection and consider a meta-analysis from both sides of the likelihood of app use to create a more homogeneous picture of the factors influencing it. For this, interdisciplinary collaboration among technology, medicine, and psychology is essential, which is why we want to encourage further work in an interdisciplinary context.

Finally, the questionnaire developed in this study can also be used with minor adaptations in other medical fields that want to work with digital apps to obtain the opinions of medical specialists.

Conclusions

The DHCQ revealed good test theoretical measures and showed signs of moderate to large correlations between the DHCQ dimensions *acceptance*, *attitude toward technology*, *technology experience*, *payment for the time of use*, and *effort of the collection* and the dependent variable *likelihood of use*. Although there were some critical voices within the group of physicians, it can be shown that there is a positive attitude and a disposition to cooperate in the development of supporting apps. It becomes clear that the likelihood of use of apps depends on more than a “good programmed app” but also requires interdisciplinary communication. The concerns about data protection and the fact that there are many apps on the market and few controls do not create an environment of confidence. However, if developers can gain the trust of physicians and mutual listening can take place to leverage the demonstrated correlations with age, experience, attitude, acceptance, effort, and payment, it is possible to work together to bridge these difficulties and enable better customization of apps to meet physicians’ needs.

Indeed, dementia is a disease that promises a better future if diagnosed early, but it is not the only one with a gradual progression and better prospects if diagnosed early. There is a great need for *good* and profitable measurement tools that are accepted by all stakeholders. However, this can only be achieved by involving physicians in the development of their working material.

Acknowledgments

The authors would like to thank all physicians who participated in the study; the Local Ethics Committee Landau for reviewing the questionnaire; and the Associations of Statutory Health Insurance Physicians (Kassenärztliche Vereinigungen) of Baden-Württemberg, Thuringia, and Saxony for distributing the study. In addition, the authors would like to thank the University of Landau and the Research Center for Information Technology (Karlsruhe); the Karlsruhe Institute of Technology (Karlsruhe); the PP.rt Clinic for Psychiatry and Psychosomatics (Reutlingen) for the opportunity to cooperate; and the Ministry of Social Affairs, Health, and Integration Baden-Württemberg, Germany, for their assistance with funding for project artificial intelligence–based platform for outpatient early detection and individualized treatment of dementia (ALFREDO).

Conflicts of Interest

None declared.

Multimedia Appendix 1
The Digital Health Compliance Questionnaire developed and evaluated in this study.
[DOCX File , 30 KB - formative_v6i6e35961_app1.docx]

Multimedia Appendix 2
Supplementary tables and figures with additional descriptive information.
[DOCX File , 290 KB - formative_v6i6e35961_app2.docx]

References

1. Bevölkerung Deutschlands bis 2060 - Tabellenband - Ergebnisse der 14. koordinierten Bevölkerungsvorausberechnung 2019 - Hauptvarianten 1 bis 9 -. Destatis. URL: https://tinyurl.com/wbdzpzmm [accessed 2021-12-24]
2. Population estimates and projections. In: Encyclopedia of the U.S. Census. Washington, D.C: CQ Press; 2012.
3. Bevölkerungsbezogene Verteilungswerte ausgewählter Laborparameter aus der Studie zur Gesundheit von Kindern und Jugendlichen in Deutschland (KiGGS). Robert Koch-Institut. URL: https://edoc.rki.de/handle/176904/3219 [accessed 2021-12-24]
4. Ziegler U, Doblhammer G. [Prevalence and incidence of dementia in Germany--a study based on data from the public sick funds in 2002]. Gesundheitswesen 2009 May;71(5):281-290. [doi: 10.1055/s-0028-1119384] [Medline: 19288422]
5. Birbaumer N, Schmitt RF. Biologische Psychologie. Cham: Springer; 2010.
6. Die Häufigkeit von Demenzerkrankungen. Deutsche Alzheimer Gesellschaft e.V. 2020. URL: https://www.deutsche-alzheimer.de/fileadmin/Alz/pdf/factsheets/infoblatt1_haeufigkeit_demenzerkrankungen_dalzg.pdf [accessed 2021-12-24]

7. Numbers of people with dementia worldwide. Alzheimer's Disease International. URL: https://www.alzint.org/resource/numbers-of-people-with-dementia-worldwide/ [accessed 2021-12-24]
8. Panegyres P, Berry R, Burchell J. Early dementia screening. Diagnostics (Basel) 2016 Jan 21;6(1):6 [FREE Full text] [doi: 10.3390/diagnostics6010006] [Medline: 26838803]
9. Dementia A Public Health Priority. Geneva: World Health Organization; 2012.
10. Die Diagnose der Alzheimer-Krankheit und anderer Demenzerkrankungen. Deutsche Alzheimer Gesellschaft. URL: https://www.deutsche-alzheimer.de/fileadmin/Alz/pdf/factsheets/infoblatt3_diagnose_dalzg.pdf [accessed 2021-12-24]
11. Dilling H, Mombour W, Schmidt MH. Internationale Klassifikation psychischer Sto¨rungen: ICD-10, Kapitel V (F) – Klinisch-diagnostische Leitlinien. Goettingen: Hogrefe; 2015.
12. Chaudhry B, Wang J, Wu S, Maglione M, Mojica W, Roth E, et al. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. Ann Intern Med 2006 May 16;144(10):742-752 [FREE Full text] [doi: 10.7326/0003-4819-144-10-200605160-00125] [Medline: 16702590]
13. Becker S, Miron-Shatz T, Schumacher N, Krocza J, Diamantidis C, Albrecht U. mHealth 2.0: experiences, possibilities, and perspectives. JMIR Mhealth Uhealth 2014 May 16;2(2):e24 [FREE Full text] [doi: 10.2196/mhealth.3328] [Medline: 25099752]
14. Grünbuch der Kommission über Mobile-Health-Dienste ("mHealth"). Bundesrat. URL: https://www.umwelt-online.de/PDFBR/2014/0167_2D14.pdf [accessed 2021-12-24]
15. Haluza D, Hofer F. Exploring perceptions on medical app use in clinical communication among Austrian physicians: results of a validation study. Health Informatics J 2020 Sep 02;26(3):1659-1671 [FREE Full text] [doi: 10.1177/1460458219888420] [Medline: 32723170]
16. Schinle M, Erler C, Schneider T, Plewnia J, Stork W. Data-driven development of digital health applications on the example of dementia screening. In: Proceedings of the 2021 IEEE International Symposium on Medical Measurements and Applications (MeMeA). 2021 Presented at: 2021 IEEE International Symposium on Medical Measurements and Applications (MeMeA); Jun 23-25, 2021; Lausanne, Switzerland. [doi: 10.1109/memea52024.2021.9478676]
17. Rutz M, Kühn D, Dierks ML. Opportunities and risks of health apps. In: Law & Net. Glashütte, Germany: Nomos; 2018.
18. Rohleder DB, Reinhardt DK. Gesundheit 4.0 – Wie Ärzte die digitale Zukunft sehen. Bitkom & Hartmannbund. URL: https://www.hartmannbund.de/wp-content/uploads/2019/08/2017_HB-Bitkom_Start-ups.pdf [accessed 2021-12-24]
19. Terry M. Medical apps for smartphones. Telemed J E Health 2010 Feb;16(1):17-22. [doi: 10.1089/tmj.2010.9999] [Medline: 20070172]
20. Ozdalga E, Ozdalga A, Ahuja N. The smartphone in medicine: a review of current and potential use among physicians and students. J Med Internet Res 2012 Sep 27;14(5):e128 [FREE Full text] [doi: 10.2196/jmir.1994] [Medline: 23017375]
21. Mosa AS, Yoo I, Sheets L. A systematic review of healthcare applications for smartphones. BMC Med Inform Decis Mak 2012 Jul 10;12:67 [FREE Full text] [doi: 10.1186/1472-6947-12-67] [Medline: 22781312]
22. Kuhn B, Amelung V. Gesundheits-Apps und besondere Herausforderungen. In: Chancen und Risiken von Gesundheits-Apps (CHARISMHA). Braunschweig: Peter L. Reichertz Institut für Medizinische Informatik der TU Braunschweig und der Medizinischen Hochschule Hannover; 2016.
23. Charani E, Castro-Sánchez E, Moore LS, Holmes A. Do smartphone applications in healthcare require a governance and legal framework? It depends on the application!. BMC Med 2014 Feb 14;12:29 [FREE Full text] [doi: 10.1186/1741-7015-12-29] [Medline: 24524344]
24. Oberman K, Müller P, Woerns S. Ärzte im Zukunftsmarkt Gesundheit 2017: Transsektorale Zusammenarbeit. Ärztinnen und Ärzte im Zukunftsmarkt Gesundheit. 2017. URL: https://www.stiftung-gesundheit.de/pdf/studien/Aerzte_im_Zukunftsmarkt_Gesundheit_2017.pdf [accessed 2021-12-24]
25. Groß D, Schmidt M. [Ethical perspectives on E-health and health apps : is all that is achievable desirable?]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2018 Mar 30;61(3):349-357. [doi: 10.1007/s00103-018-2697-z] [Medline: 29383392]
26. Livingston G, Huntley J, Sommerlad A, Ames D, Ballard C, Banerjee S, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. The Lancet 2020 Aug;396(10248):413-446. [doi: 10.1016/s0140-6736(20)30367-6]
27. Illiger K, Hupka M, von Jan U, Wichelhaus D, Albrecht U. Mobile technologies: expectancy, usage, and acceptance of clinical staff and patients at a university medical center. JMIR Mhealth Uhealth 2014 Oct 21;2(4):e42 [FREE Full text] [doi: 10.2196/mhealth.3799] [Medline: 25338094]
28. Chen J, Park Y, Putzer GJ. An examination of the components that increase acceptance of smartphones among healthcare professionals. Electron J Health Informatics 2010;5(2) [FREE Full text]
29. Park Y, Chen JV. Acceptance and adoption of the innovative use of smartphone. Industrial Manag Data Syst 2007 Nov 06;107(9):1349-1365. [doi: 10.1108/02635570710834009]
30. Putzer GJ, Park Y. Are physicians likely to adopt emerging mobile technologies? Attitudes and innovation factors affecting smartphone use in the Southeastern United States. Perspect Health Inf Manag 2012;9:1b [FREE Full text] [Medline: 22737094]
31. Wangler J, Jansky M. Gesundheits-Apps als Instrumente der Prävention? – Eine Interviewstudie zu Potenzialen für das hausärztliche Setting. Präv Gesundheitsf 2020 Mar 31;15(4):340-346. [doi: 10.1007/s11553-020-00769-x]

32. Ignatowicz A, Slowther A, Elder P, Bryce C, Hamilton K, Huxley C, et al. Ethical implications of digital communication for the patient-clinician relationship: analysis of interviews with clinicians and young adults with long term conditions (the LYNC study). BMC Med Ethics 2018 Feb 23;19(1):11 [FREE Full text] [doi: 10.1186/s12910-018-0250-0] [Medline: 29475437]
33. Sezgin E, Özkan-Yildirim S, Yildirim S. Understanding the perception towards using mHealth applications in practice. Inf Develop 2016 Dec 19;34(2):182-200. [doi: 10.1177/0266666916684180]
34. Brönneke JB, Debatin JF, Hagen J, Kircher P, Matthies H. DiGA VADEMECUM. Germany: Beck-elibrary; 2020.
35. Leitgeb N. Sicherheit von Medizingeräten. Cham: Springer; 2010.
36. Thabtah F, Peebles D, Retzler J, Hathurusingha C. A review of dementia screening tools based on mobile application. Health Technol 2020 May 09;10(5):1011-1022. [doi: 10.1007/s12553-020-00426-5]
37. Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5). Virginia, United States: American Psychiatric Association; 2018.
38. Diehl RR, Schmauss M, Berlit P. Fort- und Weiterbildung Praxisorientiertes Lernen für Neurologie und Psychiatrie - Demenz. Fortschritte der Neurologischen Psychiatrie 2003;71:617-628.
39. Umfrage Medizin- und Gesundheits-Apps. IKK Classic. 2015. URL: https://www.ikk-classic.de/assets/919_ikkc_web_pdf.pdf [accessed 2021-12-24]
40. Kramer U. Wie gut sind Gesundheits-Apps? Aktuel Ernahrungsmed 2017 Jun 20;42(03):193-205. [doi: 10.1055/s-0043-109130]
41. Urs-Vito Albrecht UV, Höhn M, von Jan U. Gesundheits-Apps und Markt. In: Chancen und Risiken von Gesundheits-Apps (CHARISMHA). Braunschweig: Peter L. Reichertz Institut für Medizinische Informatik der TU Braunschweig und der Medizinischen Hochschule Hannover; 2016.
42. Röthlisberger F, Sojer R, Zingg T, Rayki O. Die Digitalisierung aus Ärztesicht (Teil II). Schweiz Ärzteztg 2018 Nov 28:1686-1689. [doi: 10.4414/saez.2018.17377]
43. Jersak LC, da Costa AC, Callegari DA. A systematic review on mobile health care. Tecnical Report Series. 2013. URL: https://www.pucrs.br/facin-prov/wp-content/uploads/sites/19/2016/03/tr073.pdf [accessed 2021-12-24]
44. Jahresbericht 2020. Ärztekammer Nordrhein. URL: https://www.aekno.de/fileadmin/user_upload/aekno/downloads/2020/jahresbericht-2020.pdf [accessed 2022-05-18]
45. Albrecht UV. Chancen und Risiken von Gesundheits-Apps. In: Recht & Netz. Medizinische Hochschule Hannover: Nomos; 2018.
46. Schinle M, Wyszka D, Schwärzler F, Volz K, Ruby MA, Sejdinovic E, et al. An approach to digitalize psychological tests to support diagnosis of Alzheimer's disease in ambulatory care. In: Proceedings of the 2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA). 2018 Presented at: 2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA); Jun 11-13, 2018; Rome, Italy. [doi: 10.1109/memea.2018.8438761]
47. Rosseel Y. lavaan: an R package for structural equation modeling. J Stat Softw 2012;48(2):1-36. [doi: 10.18637/jss.v048.i02]
48. Epskamp S. semPlot: path diagrams and visual analysis of various SEM packages' output. CRAN. 2019. URL: https://cran.r-project.org/web/packages/semPlot/index.html [accessed 2021-12-24]
49. Wickham H. Data analysis. In: ggplot2. Use R!. Cham: Springer; 2016.
50. Lachenbruch PA, Cohen J. Statistical power analysis for the behavioral sciences (2nd ed.). J Am Stat Assoc 1989 Dec;84(408):1096. [doi: 10.2307/2290095]
51. Janneck M, Vincent-Höper S, Othersen I. Entwicklung und Validierung eines Fragebogens zum Technikbezogenen Selbstkonzept (TSK): eine gendersensitive studie. Gruppendyn Organisationsberat 2012 Jul 4;43(3):289-310. [doi: 10.1007/s11612-012-0184-9]
52. Karrer K, Glaser C, Clemens C, Bruder C. Technikaffinität erfassen - Der Fragebogen TA-EG. Klinikarzt 2009 Apr 30;38(04):197-199. [doi: 10.1055/s-0029-1223266]
53. Schwärzler F, Roland P. PflegeCoDe - Digitalization of Psychological Tests to Support AD Diagnosis: challenges and Opportunities. PflegeCoDe Abschlussbericht (Zeitraum 2015 - 2018). 2018. URL: https://www.tib.eu/de/suchen/id/TIBKAT:1663697639/Titel-des-Vorhabens-PflegeCoDe-PflegeCoDe-Abschlussbericht [accessed 2021-12-24]
54. Kim HY. Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. Restor Dent Endod 2013 Feb;38(1):52-54 [FREE Full text] [doi: 10.5395/rde.2013.38.1.52] [Medline: 23495371]
55. Backhoff E, Larrazolo N, Rosas M. The level of difficulty and discrimination power of the Basic Knowledge and Skills Examination (EXHCOBA). Revista Electrónica de Investigación Educativa 2000;2(1).
56. Nemoto T, Beglar D. Developing Likert-scale questionnaires. In: Proceedings of the JALT 2013 Conference. 2013 Presented at: JALT 2013 Conference Proceedings; Oct 25-28, 2013; Tokyo.
57. D'Sa JL, Visbal-Dionaldo ML. Analysis of multiple choice questions: item difficulty, discrimination index and distractor efficiency. Int J Nurs Educ 2017;9(3):109. [doi: 10.5958/0974-9357.2017.00079.4]
58. Eid M, Gollwitzer M, Schmitt M. Statistik und Forschungsmethoden: Lehrbuch. Weinheim, Germany: Psychologie Verlags Union; 2015.
59. Messer M, Schneider G. Statistik Theorie und Praxis im Dialog. Cham: Springer; 2019.

60. Mîndrilă D. Maximum Likelihood (ML) and Diagonally Weighted Least Squares (DWLS) estimation procedures: a comparison of estimation bias with ordinal and multivariate non-normal data. Int J Digital Soc 2010 Mar 1;1(1):60-66. [doi: 10.20533/ijds.2040.2570.2010.0010]
61. Taasoobshirazi G, Wang S. The performance of the SRMR, RMSEA, CFI, and TLI: an examination of sample size, path size, and degrees of freedom. J Applied Quant Methods 2016;11(3):31.
62. Said H, Badru BB, Shahid M. Confirmatory Factor Analysis (Cfa) for testing validity and reliability instrument in the study of education. Aus J Basic Applied Sci 2011;5(12):1098-1103.
63. Atkinson TM, Rosenfeld BD, Sit L, Mendoza TR, Fruscione M, Lavene D, et al. Using confirmatory factor analysis to evaluate construct validity of the Brief Pain Inventory (BPI). J Pain Symptom Manag 2011 Mar;41(3):558-565 [FREE Full text] [doi: 10.1016/j.jpainsymman.2010.05.008] [Medline: 21131166]
64. Straub D, Gefen D. Validation guidelines for IS positivist research. Commun Assoc Inf Syst 2004;13. [doi: 10.17705/1CAIS.01324]
65. Streiner DL. Starting at the beginning: an introduction to coefficient alpha and internal consistency. J Personality Assessment 2003 Feb;80(1):99-103. [doi: 10.1207/s15327752jpa8001_18]
66. Tavakol M, Dennick R. Making sense of Cronbach's alpha. Int J Med Educ 2011 Jun 27;2:53-55 [FREE Full text] [doi: 10.5116/ijme.4dfb.8dfd] [Medline: 28029643]
67. Parry S. Fit indices commonly reported for CFA and SEM. Cornell University. URL: http:///C:/Users/user/Downloads/Cornell_SEM_fit%20(1).pdf [accessed 2021-12-24]
68. Comrey AL, Lee HB. A First Course in Factor Analysis. Milton Park, Abingdon-on-Thames, Oxfordshire United Kingdom: Taylor & Francis; 1973.
69. Carter SR. Using confirmatory factor analysis to manage discriminant validity issues in social pharmacy research. Int J Clin Pharm 2016 Jun 4;38(3):731-737. [doi: 10.1007/s11096-016-0302-9] [Medline: 27147255]
70. mhGAP Intervention Guide - Version 2.0 for Mental, Neurological and Substance Use Disorders in Non-specialized Health Settings. Geneva: World Health Organization; 2019.

Abbreviations

CFA: confirmatory factor analysis
CFI: comparative fit index
DHCQ: Digital Health Compliance Questionnaire
DWLS: diagonally weighted least squares
TLI: Tucker-Lewis index

Edited by G Eysenbach; submitted 24.12.21; peer-reviewed by A Ren, A Garcia; comments to author 03.04.22; revised version received 08.04.22; accepted 29.04.22; published 22.06.22.

Please cite as:
Schinle M, Erler C, Kaliciak M, Milde C, Stock S, Gerdes M, Stork W
Digital Health Apps in the Context of Dementia: Questionnaire Study to Assess the Likelihood of Use Among Physicians
JMIR Form Res 2022;6(6):e35961
URL: https://formative.jmir.org/2022/6/e35961
doi:10.2196/35961
PMID:35731567

Original Paper

Improving Pain Assessment Using Vital Signs and Pain Medication for Patients With Sickle Cell Disease: Retrospective Study

Swati Padhee[1], MTech, MS; Gary K Nave Jr[2], PhD; Tanvi Banerjee[1], PhD; Daniel M Abrams[2], PhD; Nirmish Shah[3], MD

[1]Department of Computer Science and Engineering, Wright State University, Dayton, OH, United States

[2]Department of Engineering Sciences and Applied Mathematics, Northwestern University, Chicago, IL, United States

[3]Division of Hematology, Duke University School of Medicine, Durham, NC, United States

Corresponding Author:
Swati Padhee, MTech, MS
Department of Computer Science and Engineering, Wright State University
3640 Colonel Glenn Hwy
Dayton, OH, 45435
United States
Phone: 1 937 775 1000
Email: padhee.2@wright.edu

Abstract

Background: Sickle cell disease (SCD) is the most common inherited blood disorder affecting millions of people worldwide. Most patients with SCD experience repeated, unpredictable episodes of severe pain. These pain episodes are the leading cause of emergency department visits among patients with SCD and may last for several weeks. Arguably, the most challenging aspect of treating pain episodes in SCD is assessing and interpreting a patient's pain intensity level.

Objective: This study aims to learn deep feature representations of subjective pain trajectories using objective physiological signals collected from electronic health records.

Methods: This study used electronic health record data collected from 496 Duke University Medical Center participants over 5 consecutive years. Each record contained measures for 6 vital signs and the patient's self-reported pain score, with an ordinal range from 0 (no pain) to 10 (severe and unbearable pain). We also extracted 3 features related to medication: *medication type*, *medication status (given or applied*, or *missed or removed or due)*, and *total medication dosage (mg/mL)*. We used variational autoencoders for representation learning and designed machine learning classification algorithms to build pain prediction models. We evaluated our results using an accuracy and confusion matrix and visualized the qualitative data representations.

Results: We designed a classification model using raw data and deep representational learning to predict subjective pain scores with average accuracies of 82.8%, 70.6%, 49.3%, and 47.4% for 2-point, 4-point, 6-point, and 11-point pain ratings, respectively. We observed that random forest classification models trained on deep represented features outperformed models trained on unrepresented data for all pain rating scales. We observed that at varying Likert scales, our models performed better when provided with medication data along with vital signs data. We visualized the data representations to understand the underlying latent representations, indicating neighboring representations for similar pain scores with a higher resolution of pain ratings.

Conclusions: Our results demonstrate that medication information (the type of medication, total medication dosage, and whether the medication was given or missed) can significantly improve subjective pain prediction modeling compared with modeling with only vital signs. This study shows promise in data-driven estimated pain scores that will help clinicians with additional information about the patient's condition, in addition to the patient's self-reported pain scores.

(JMIR Form Res 2022;6(6):e36998) doi:10.2196/36998

KEYWORDS

pain management; pain medication; vital signs; sickle cell disease; machine learning

Introduction

Background

Sickle cell disease (SCD) is a family of genetic blood disorders that affects >20 million people worldwide [1], the most prevalent complication of which is pain. Pain crises in SCD are strongly linked to increased medical costs, morbidity, and mortality [2]. During childhood, SCD often presents as unpredictable, severe, acute pain episodes characterized by pain periods ranging from hours to weeks, which usually occur a few times a year. The most challenging aspect of treating pain episodes in SCD is the assessment and interpretation of the patient's pain intensity level [3,4]. However, in current clinical practice, patient self-reporting is the gold standard approach to determining the absence, presence, and severity of pain [4,5].

Furthermore, because of the subjective nature of pain, it is challenging for clinicians to precisely ascertain the severity of the patient's pain. This assessment is particularly difficult in patients with chronic pain. Furthermore, effective treatment strategies for patients with SCD, such as intravenous opioid therapy, are palliative. Ultimately, as pain is inherently subjective, medical providers and patients have difficulty in determining the ideal treatment and management strategies for pain. As a result, there has been an increasing focus on developing and implementing pain prediction models from objective measures over the past several years [3,4,6-8]. However, in addition to the slow development of these models, the difficulty also lies in understanding the severity of a patient's pain level and their response to pain management strategies.

Currently, the standard treatment protocol for painful episodes associated with SCD includes rest, aggressive hydration, treatment of any underlying infections or other complications, and a focus on analgesics such as opioids [9,10]. However, there is wide variability in the management of painful episodes in hospitals. Variations in practice reflect different views about the suitability of opioids, such as concerns of dependence on opioids. In addition, each patient has historical differences in responses to opioids; the methods of administration of opioids, such as continuous infusion and patient-controlled analgesia, lead to varied responses in patients. Pharmacological management strategies for acute pain associated with SCD include opioids, nonopioids, and adjuvant analgesics or coanalgesics. Medication strategies for chronic pain are diverse and lead to substantial variability. Hence, it is essential to consider the inclusion of pain medication when modeling pain prediction.

The current literature shows an increasing focus on machine learning (ML) techniques to understand the various complexities associated with patient health in SCD [11-14]. Lazakidou et al [15] developed a personal electronic health record (EHR) to evaluate the deployment of an advanced web-based application platform that assessed health care professionals and patients to provide a more efficient and effective solution than that of the daily clinical routine. In their study, web-based solutions enabled patients to update and access their medical information. The system was examined with 3 varied patient groups comprising 150 patients with Parkinson disease, diabetes, and congenital heart disease engaged in 3 European clinics. The outcomes indicated that personal EHRs could provide better services in terms of user-friendliness, data management, comprehensiveness, and valuable content. Du et al [16] developed a microfluidic device that could examine the behavior of blood from patients with SCD. This device could also measure how long blood cells took to become stiff and get stuck in the blood vessels. A total of 25 patients with SCD were involved in their study. By using this device to evaluate blood samples, the researchers were able to decide how deoxygenation affected the sickling rates of red blood cells (RBCs), capillary' stick rates, and how quickly the RBCs reshaped, especially when oxygen levels were restored. Knowlton et al [17] presented a sensitive, label-free, and specific testing platform to diagnose SCD using blood samples based on the density of sickle RBCs under deoxygenated conditions. Using this platform, they could differentiate between the levitation patterns of sickle and control RBCs in association with their degree of confinement.

In the face of the continued opioid crisis, the search for more objective measures of pain continues to evolve rapidly in medicine, and studies examining a variety of objective measures to predict pain have been published in recent years [7,18]. Prior studies have reported preliminary evidence that fluctuations in vital signs may be used to assess pain in patients in the intensive care unit [19] as acute pain leads to changes in vital signs [20]. These physiological measures include blood pressure, respiratory rate, oxygen saturation, temperature, and pulse rate. Nickerson et al [21] predicted pain scores—measured between 40 and 120 minutes after administering 10 mg of oxycodone—from pain score values before drug administration using 200 features for each patient in the electronic medical records data. Essential features included age, gender, Charlson comorbidity index, BMI, ethnicity, and International Classification of Diseases ninth edition code class. They predicted a postmedication (oxycodone) pain score with an accuracy of 66% using support vector regression. They concluded that these results would likely improve with more temporal data (eg, vital signs), which we explore in this study by using both vital signs and medication data to predict the severity of pain with multiple Likert scales, with the best performance of 82.8% accuracy using only 9 features, as shown in Textbox 1. Although most prior studies have explored this question in the context of misuse of pain medications (particularly regarding abuse of opioid medications), we used the pain medication provided to patients to predict the severity of pain. The ability to objectively and accurately predict pain severity and onset could result in more prompt and effective treatment of pain crises, leading to improved outcomes and encouraging more diligent use of medications [22]. Although there are complications associated with using medication data for prediction at the same time as pain measurements, we investigated their application here to provide a baseline for this comparison and open the door for future research involving dynamic pain and medication measurements. Our principal hypothesis was that pain medications help with better pain-related function and pain intensity management.

Several previous studies from our research group used EHR data to predict pain. Yang et al [4] initially demonstrated the

feasibility of ML techniques on a limited data set of 5363 records from 40 patients during inpatient hospital visits to predict subjective pain scores from 6 objective vital signs with support vector machines (SVMs), achieving an accuracy of 58.2%. Alambo et al [8] examined 424 clinical notes from the same cohort of 40 patients to predict the prevalence of pain and whether pain increased, decreased, or remained constant. Padhee et al [6], using 6 objective vital signs and the nature of hospital visit information from 59,728 records of a different cohort of 47 patients over 5 consecutive years, demonstrated that with more data for each patient, the accuracy for pain prediction improved (accuracy of 65.3%). In this study, we demonstrated that with more data from a larger cohort of patients and medication information, the accuracy improved by 17.5% compared with Padhee et al [6] and by 24.6% compared with Yang et al [4].

Deep neural networks have been shown to contribute promising capabilities in learning complex patterns in data and have achieved remarkable success in several domains such as computer vision, natural language processing, and speech recognition. Recently, many efforts have been made to improve the performance of ML tasks in the field of biomedical and health informatics. Deep learning has previously been successfully used on EHRs to achieve both specific and general goals [23]; for instance, both *Deep Patient* [24] and *Doctor AI* [25] used unsupervised deep learning before supervised learning. As in many other applications, the challenge of missing data is common in ML studies applied to EHR data, which often contain entries with missing elements. These challenges arise as the data are manually collected from patients and may vary depending on circumstances. In this study, we address this challenge using variational autoencoders (VAEs) [26,27] to reconstruct missing data. VAEs are unsupervised deep feature methods that provide data reconstruction by probabilistically filling in the data between the encoding and decoding steps. As the encoder neural network typically expects a fixed-length vector as input, the question arises regarding what we can do with the missing values in the VAE encoder input. We followed the heuristic of replacing missing elements with fixed values [28,29]. Although VAEs have been previously applied to EHR data [30], we show here that they improve pain prediction capabilities from physiological signs with and without medication information.

Textbox 1. Data modalities and variables considered in this study.

Vital signs

- Peripheral capillary oxygen saturation
- Systolic blood pressure
- Diastolic blood pressure
- Heart rate
- Respiratory rate
- Temperature

Medication

- Medication type (5 classes)
  - Hydromorphone
  - Acetaminophen
  - Ketorolac
  - Oxycodone
  - Fentanyl
- Medication status (2 classes)
  - Given or applied
  - Missed or removed or due
- Total medication dosage (mg/mL)

Pain

- Self-reported pain score on a scale of 0-10 (0=no pain to 10=severe and unbearable pain)

Objective

This study aimed to use vital signs and medication information collected from the EHR data of patients with SCD to predict patient-reported pain scores using ML techniques. In this paper, we propose to represent multiple data modalities in EHRs in high-level abstraction, vital signs, and medication information using deep autoencoder networks such as VAEs to predict pain intensity on varying Likert scales. Our specific contributions are as follows:

1. To the best of our knowledge, we analyzed the most extensive EHR data of 126,519 records from 496 patients with SCD collected over 5 consecutive years and demonstrated that a larger patient cohort data improves model performance in pain prediction. We reduced the data to 33,000 records by removing data with multiple medication fields missing and then used them for our model evaluation.
2. We showed that pain medication information with vital signs data can improve pain prediction at varying pain rating scales (ie, different granularities).
3. We demonstrated that deep representational learning can not only improve pain prediction results but also provide a better understanding of the role of medication and physiology on the patient's pain response with a patient profiles study.

Methods

Data

In this study, we analyzed inpatient and outpatient EHR data collected from 496 participants at Duke University Hospital over 5 consecutive years. Each record contained measures of 6 vital signs, as shown in Textbox 1. In addition to the vital signs, each record also included the patient's self-reported pain score, with an ordinal range from 0 (no pain) to 10 (severe and unbearable pain). The pain score was recorded by the medical staff during outpatient visits when the patients reported no pain (pain score 0) and during their monitoring of inpatient visits. The vital signs were recorded by the medical staff every 4 hours for inpatient stays, and the medication data were recorded as given to the patients. We extracted 3 medicinal features from the data upon consultation with our coauthor physician, as shown in Textbox 1. We calculated the total medication dosage as the sum of all medication dosages recorded for a patient at a given time t using the following equation:

Here, $Medication\ Dosage_i(t)$ indicates the dosage of ith medication type recorded at time t for the patient.

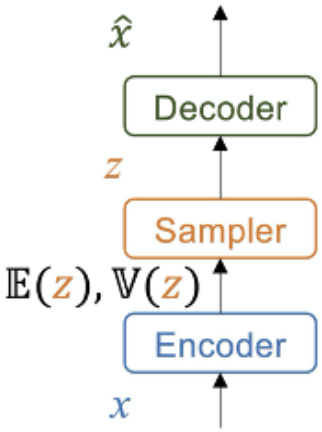
We removed the data points in this study for which the Medication Administration Record was on hold.

Background

In this study, we used VAEs to impute missing values within the data based on other samples. Autoencoders are a class of unsupervised deep learning techniques in which neural networks are leveraged for the task of representation learning. We designed a neural network architecture to impose a bottleneck in the network, thereby forcing a compressed knowledge representation of the original input data modalities. If the input features were such that they were independent of one another, this compression and subsequent reconstruction would be an arduous task. However, if some association exists in the data (eg, correlations between input data modalities), this structure can be learned and consequently leveraged when forcing the input through the network's bottleneck. VAEs are probabilistic generative models that have the same architecture as vanilla autoencoders but consider specific assumptions about the distribution of middle or latent layer variables. They learn the true distribution of input features from latent variable distributions using a Bayesian approach and present a theoretical framework for reconstruction and regularization [31].

A VAE learns the distribution of data with an encoder network by fitting it to a Gaussian distribution and generates data with a decoder by sampling from the learned distribution. We used autoencoders to reconstruct the input data (x) in the output  layer by an encoding and decoding process. As shown in Figure 1, the encoder network converts the input data (x) into a latent representation (z). The hidden state comprises 2 additional layers: $E(z)$ and $V(z)$, where the latent variable z follows a Gaussian distribution with mean $E(z)$ and variance $V(z)$. We sample z from the distribution parameterized by the encoder; the decoder network then remodels the input from the latent representations by using z to generate  . The fundamental property of autoencoders is that they minimize this reconstruction error using a loss function comprising a reconstruction term ($l_{reconstruction}$), which is the mean squared error between the output and the input, and a regularization term (Kullback-Leibler divergence loss [l_{KL}]), as shown in the following equation:

Figure 1. An illustration of the variational autoencoder architecture for one input data modality.



The term  is on the final layer and the regularization term enforces a specific Gaussian structure on the latent layer through a penalty term $l_{KL}(z, N(0, I_d))$. β in the loss function is a hyperparameter that dictates how to weigh the reconstruction and penalty terms.

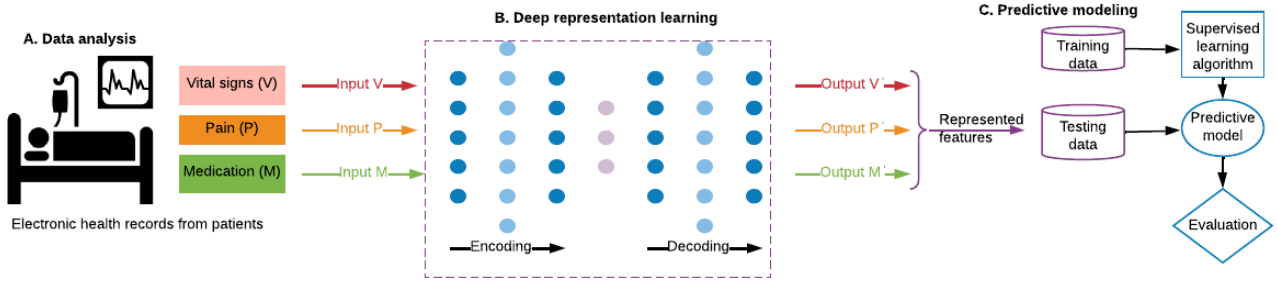
Variation in VAE means that the encoder network estimates the μ (mean) and σ (SD) parameters (latent variables) of the Gaussian distribution. However, real-world applications, including health care, almost always have missing values. In correspondence with the missing values in the raw temporal data, we substituted the corresponding categories with a unique integer to properly encode the status of the missing information. The encoder comprises a long short-term memory cell. It receives input sequences resulting from the concatenation of the raw physiological data and the extracted categorical medicinal features. As in every encoder in a VAE architecture, it produces an output that is used to approximate the mean and variance of the latent distribution. The decoder samples from the latent distribution form the output sequences. This approach helps us develop an unsupervised framework that can fill the missing pieces appearing in real-world EHR data volume streams, not only in patients with SCD but also in other health care applications.

Proposed Framework

Overview

Figure 2 provides an overview of the proposed approach in 3 consecutive steps. In step A, we preprocessed the raw data to overcome data challenges such as missing values. Next, in step B, we applied unsupervised deep representation learning to generate higher-level abstraction of the input data modalities. Finally, in step C, we investigated supervised algorithms for predictive modeling and performed the evaluation.

Figure 2. Deep representation learning for pain prediction (A: data analysis, B: deep representation learning, and C: predictive modeling).



Step A: Data Analysis

In this study, we used records from EHR data collected at Duke University Hospital and identified them using study labels to label patients without identification. The timestamp for each data entry was deidentified, preserving temporality. The data set had missing values for ≥1 vital sign, medication, and pain score. The data set contained 126,519 records from 496 patients collected over 5 consecutive years. However, we included 33,000 records in this study owing to >4 missing features in the remaining records. Of the 33,000 records, 18,291 (55.43%) included at least 1 of the 5 administered medication types (Textbox 1). The demographic information of the patients was not available. Data for each patient varied; although 70 patients had a one-time visit to the hospital, 240 patients visited for at least >100 days. Most patient records were for a patient staying for 1705 days with a high mean pain score of 8 who received pain medication 219 times (an average of 338 mg of total pain medication dosage). We did not consider the effect of any other medical condition on the patients in this study.

Step B: Deep Representation Learning

In the second step, we represented all input data modalities in high-level abstraction using multiple deep autoencoder networks, including VAEs. We evaluated the performance of each network while considering the tuning of hyperparameters such as the learning rate, batch size, number of epochs, and number of hidden layers and hidden units for training to avoid overfitting.

Step C: Predictive Modeling

In this step, we applied supervised learning techniques to the represented data set using linear and nonlinear approaches such as random forest (RF) [32], Lasso regression [33], and SVM [34]. Our experiments comprised three main phases: (1) training the VAE, where the training samples were used to train the VAE and the reconstruction loss for each training data sample was stored according to the target pain score; (2) generating new pain scores, where the VAE decoder generated new pain score samples based on specified classes and each newly generated data sample was merged into the original training data set under the condition that the class reconstruction loss was satisfied; and (3) predicting pain scores, where the VAE decoder was used to initialize the weight of the hidden layers, the merged training data set was used to train the classifier, and the trained classifier was used to predict pain scores on the testing data set.

Experimental Study

In our experimental study, we implemented our methodology on a deidentified EHR data set. This study design helped us discover our method's performance in predictive modeling for patients with SCD. Across several attributes comprising patient clinical records and individual health status, 9 attributes, including vital signs and medication information, were considered for the data analysis of 496 patients. As mentioned previously, the goal was to predict pain scores based on high-dimensional features.

We implemented the VAE using the *PyTorch* and *Keras* libraries with a *TensorFlow* backend in Python. The VAE architecture has 5 hidden layers (2 hidden layers of encoders and decoders and 1 middle layer). We applied hyperparameter tuning for major hyperparameters such as the learning rate, activation functions, and batch size to select the best hyperparameters. We used a hidden dropout component with a dropout rate of 0.2 and a sigmoid activation function for the final layers. The models were trained for 100 epochs using an Adam optimizer with a learning rate of 0.001 (with exponential decay rates of first- and second-moment estimates β_1=.9 and β_2=.999) and a batch size of 64. Once the latent features were extracted, they were fed into a supervised learning model for pain score prediction. For the supervised learning step, we considered 3 well-known supervised classifiers: RF (with 50 trees and half of the features considered at every split), Lasso regression, and SVM (with radial basis function kernel C 1.5 and gamma set to $1/N_f$, where N_f denotes the number of features). We used the grid search method to determine the optimal hyperparameters for supervised classifiers. We used the average accuracy as our evaluation measure for performance evaluation in the testing process. Finally, we visually inspected the learned representations of the entire data set and compared them with the represented data. We used t-distributed stochastic neighboring embedding (t-SNE) [35] for this task.

Ethics Approval

The study protocol was approved by the institutional review board of Duke University Medical Center in May 2018 with IRB number Pro00068979. Identifiable personal information was not collected; all data were kept confidential and safe according to the internal data security policy, and they were only accessible to authorized researchers.

Results

Vital Signs

We evaluated our approach using VAE data (represented data) and original data (unrepresented data) on supervised classifiers for pain prediction tasks and compared their performance based on the results obtained from the testing process with 5-folds cross-validation (for each fold, we considered 80% of the data for training, 10% for the validation set, and 10% for the test set). This comparison is presented in Table 1. We would like to note here that missing values for pain scores were not imputed. In our data set, we had 11 unique self-reported pain scores where patients described their experienced pain intensity on a scale of 0 to 10. It is challenging for one person to distinguish between such broad and granular pain intensity levels and be consistent in the reporting of every pain episode. Hence, in addition to the 11 pain scores, we evaluated our pain prediction models by transforming our data set into a 6-point rating scale, a 4-point rating scale, and a binary rating scale according to the following transformation rules:

1. The 6 pain scores: none=0, very mild=1 to 2, mild=3 to 4, moderate=5 to 6, severe=7 to 8, and very severe=9 to 10
2. The 4 pain scores: none=0, mild=1 to 3, moderate=4 to 6, and severe=7 to 10
3. The 2 pain scores: no or mild pain=0 to 5 and severe pain=6 to 10

As shown in Table 1, an RF-supervised classifier trained on data represented using VAE performed best in each pain rating scale, achieving the highest accuracy of 60.3% in predicting 2 pain scores (no or mild pain and severe pain). According to these results, our approach with representation learning reduces the prediction error and achieves better accuracy than using the original features.

Table 1. Pain prediction results in varying pain scales on vital signs data (accuracy) arranged from higher resolution to lower resolution.

Approach	11 pain scores			6 pain scores			4 pain scores			2 pain scores		
	RF[a]	SVM[b]	Lasso	RF	SVM	Lasso	RF	SVM	Lasso	RF	SVM	Lasso
Original data	0.301	0.242	0.216	0.363	0.352	0.337	0.432	0.426	0.392	0.535	0.513	0.486
VAE[c] data	0.343	0.321	0.307	0.391	0.371	0.348	0.472	0.452	0.439	0.603	0.561	0.549

[a]RF: random forest.

[b]SVM: support vector machine.

[c]VAE: variational autoencoder.

Vital Signs and Medicinal Data

We also analyzed the performance of pain score prediction using only vital signs compared with including medication information. We show in Table 2 that our approach with the RF classifier achieves better accuracy with medication and vital signs information than with only vital signs information in predicting the respective pain scores. This indicates that when provided with additional medication information, our approach can learn better representations of patient profiles from vital signs to predict their pain levels. The higher accuracy associated with the narrow scales is attributed to the narrow space to misclassify many records by our models, thereby improving the chances of correctly predicting the pain score.

We also show in Table 3 the area under the curve (AUC) for the receiver operating characteristic for the best-performing, clinically relevant (as suggested by our coauthor clinical partner) models (models a, b, d, and e from Table 2). Overall, the AUC for both the 2 pain score and 4 pain score rating scales suggested no discrimination. This indicates that our models can predict pain in patients based on their vital signs and medication information at various intensity levels. For the 2 pain score rating scales, an AUC of 0.92 suggests a 92% chance that our model correctly distinguishes a pain score in the no or mild pain

range (0-5 pain score) from the severe pain range (6-10 pain score) based on the patient's vital signs and medication information instead of a random assignment probability of 50%. Empirically, our results demonstrate that (1) medical feature representation can improve prediction performance, and (2) medication information can lead to significant improvement in pain level prediction.

Table 2. Pain prediction results in varying pain scales on vital signs data (accuracy) as compared with additional medication data arranged from higher resolution to lower resolution.

Approach	11 pain scores		6 pain scores		4 pain scores		2 pain scores	
	Vitals	Vitals+medicinal	Vitals	Vitals+medicinal	Vitals	Vitals+medicinal	Vitals	Vitals+medicinal
Original data	0.301	0.442	0.363	0.463	0.432	0.689[a]	0.535	0.787[b]
VAE[c] data	0.343	0.476	0.391	0.493	0.472	0.706[d]	0.603	0.828[e]

[a]Model with original representations using both vital and medication data for the 4 pain score rating scale.

[b]Model with original representations using both vital and medication data for the 2 pain score rating scale.

[c]VAE: variational autoencoder.

[d]Model with deep representations using both vital and medication data for the 4 pain score rating scale.

[e]Best-performing model with deep representations using both vital and medication data for the 2 pain score rating scale.

Table 3. Area under the curve for the receiver operating characteristic for the best-performing models (models a, b, d, and e).

Approach	4 pain scores				2 pain scores	
	None: 0	Mild: 1-3	Moderate: 4-6	Severe: 7-10	No or mild pain: 0-5	Severe pain: 6-10
Original data	0.82	0.85	0.88	0.83	0.91	0.89
VAE[a] data	0.82	0.86	0.89	0.83	0.92	0.89

[a]VAE: variational autoencoder.

Discussion

Principal Findings

Overview

Our study demonstrates that although there are complications associated with using medication data for prediction at the same time as pain measurements, ML models can be used for dynamic pain and medication measurements. Our findings indicate the importance of medication information (achieving an accuracy of 82.3%) and demonstrate that a larger cohort of patient data with deep representational learning improves model performance (by 17.5% as compared with Padhee et al [6] and by 24.6% as compared with Yang et al [4]). Furthermore, from our unsupervised analysis, we distinguished unique patient profiles (Table 4) that can help isolate different patient profiles to further understand the role of physiology and medication in pain response. In addition, there are 2 main types of opioids: short-acting analgesics and sustained-release analgesics, and the dosing pattern differs depending on the properties of these drugs. Our initial results show that considering medication type (Textbox 1), status, and dosage can improve pain assessment models, providing evidence for future studies to further analyze the variability in dosing patterns.

Table 4. Sample of patient profiles from the learned variational autoencoder representations clustered together using t-distributed stochastic neighboring embedding projections, as shown in Figure 3.

Medication administered	Region	Patient number	Correlation between medication and pain score	Vital signs	Correlation between medication dosage and vital signs
Oxycodone	1[a]	1	0.23	Temperature	0.65
Hydromorphone	1	1	0.11	Temperature	0.65
Acetaminophen	1	1	0.17	Temperature	0.65
Ketorolac	1	2	0.55	Systolic blood pressure	0.47
Hydromorphone	1	2	0.24	Systolic blood pressure	0.47
Hydromorphone	2[b]	3	0.35	Systolic blood pressure	0.099
Acetaminophen	2	3	–0.20	Systolic blood pressure	0.099
Ketorolac	2	3	0.41	Pulse	0.04
Oxycodone	3[b]	4	0.59	Systolic blood pressure	0.27
Hydromorphone	3	4	0.08	Systolic blood pressure	0.27
Fentanyl	4[c]	5	—[d]	Peripheral capillary oxygen saturation level	0.13
Acetaminophen	4	6	—	Temperature	0.47
Acetaminophen	4	6	—	Pulse	0.38
Acetaminophen	4	6	—	Peripheral capillary oxygen saturation level	0.28

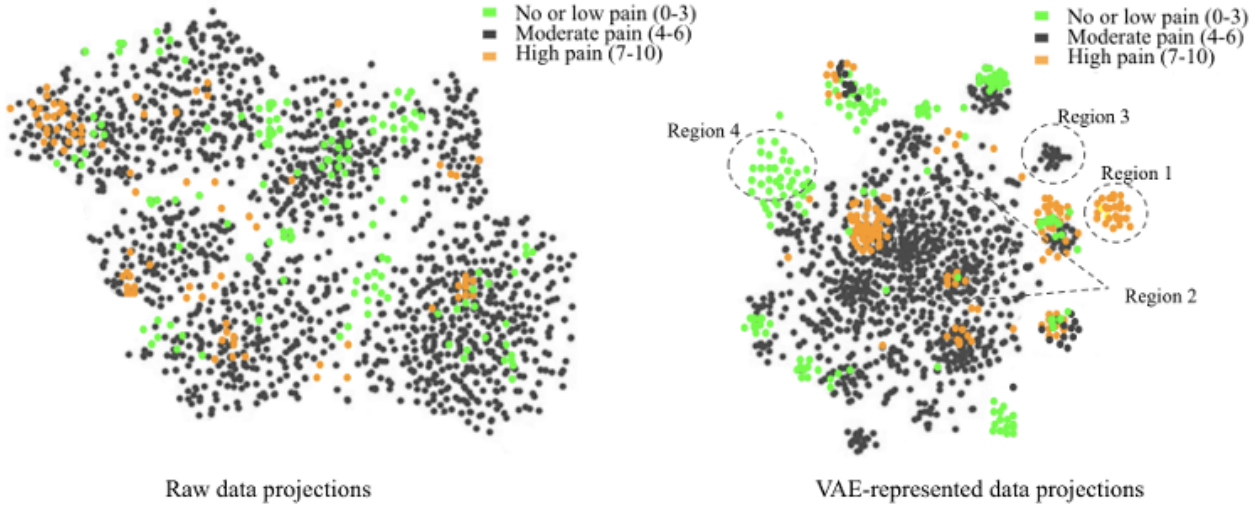
[a]High pain.

[b]Moderate pain.

[c]No or low pain.

[d]Medication not available.

Figure 3. Visualization of the learned data representations using t-SNE: t-distributed stochastic neighboring embedding projections. VAE: variational autoencoder.



Deep Representation Learning

In this study, we applied a deep feature representation to predict the pain scores of patients with SCD based on their vital signs and medication information. These results emphasize that representation learning can play an effective role in the performance of clinical prediction. As shown in Table 1, our models trained on deep represented features can identify pain scores for 6.8% more patients at an abstract pain intensity level of no or mild pain or severe pain. They also display significant improvement by detecting pain intensity for 4.2% more patients at a highly granular pain score intensity (ie, on 11 pain ratings) than models trained on unrepresented raw vital signs data. We observed a similar performance of deep represented features compared with raw data features when medicinal data were included in the modeling, which may indicate that the medication information can allow the use of simpler features by providing more pain-related information than the more

convoluted deep represented features. However, our models trained on VAE-represented features generated using both vitals and medicinal data could identify pain scores for 3.4%, 3%, 1.7%, and 4.1% more patients from higher to lower resolutions of pain intensity than models trained on raw data. To investigate further, we show the confusion matrices of the best-performing models trained on vitals and medicinal data in Figures 4-7. As shown in Figures 4 and 5, it is interesting to note that with deep feature representations, our model can accurately identify not just 87 more cases of no or mild pain but 55 more cases of severe pain while reducing misclassification. This is important to consider in a clinical setting while deciding on the diligent use of medications in a larger patient cohort.

Similarly, Figures 6 and 7 show that with more granular 4-point pain intensity levels (pain scores: none=0, mild=1-3, moderate=4-6, and severe=7-10), our model trained on deep represented features can identify more instances for each category accurately than the original data representations while reducing the misclassification. The model can identify more instances of moderate pain than none, mild, or severe pain. It is noteworthy that the misclassification for each pain category reduced with the stretch between the pain severity levels. For example, as shown in Figure 7, our best model for 4 pain scores (Table 2) incorrectly predicted 21 instances of severe pain data as no pain, 35 instances as mild pain, and 132 instances as moderate pain, highlighting that the error primarily lies in the prediction of moderate pain as no pain. Similarly, it predicted 16 instances of low pain as severe pain, 39 as moderate pain, and 141 as mild pain. Misclassification reduces with the granularity of pain intensity, reflecting the subjective nature of pain.

Figure 4. Confusion matrix for the best-performing model with original data representations for 2 pain score levels (pain scores: no or mild=0-5 and severe=6-10; Table 2, model b).

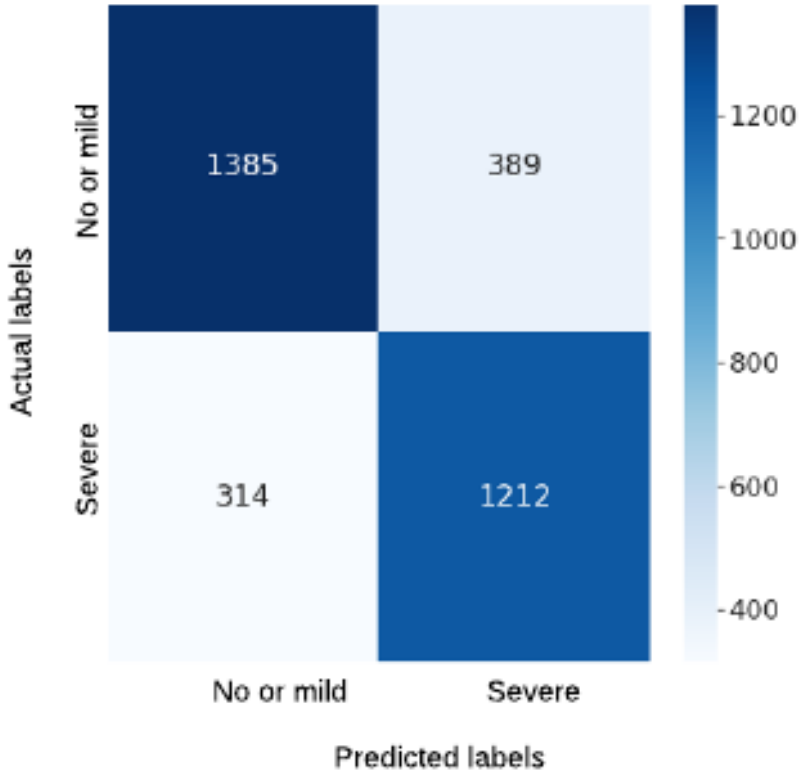


Figure 5. Confusion matrix for the best-performing model with variational autoencoder data representations for 2 pain score levels (pain scores: no or mild=0-5 and severe=6-10; Table 2, model e).

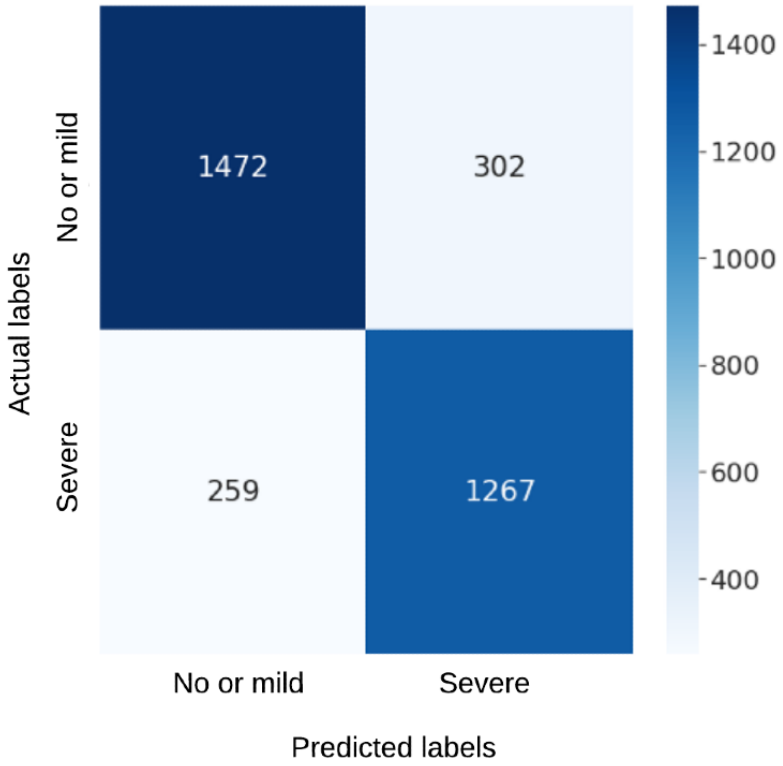


Figure 6. Confusion matrix for the best-performing model with original data representations for 4 pain score levels (pain scores: none=0, mild=1-3, moderate=4-6, and severe=7-10; Table 2, model a).

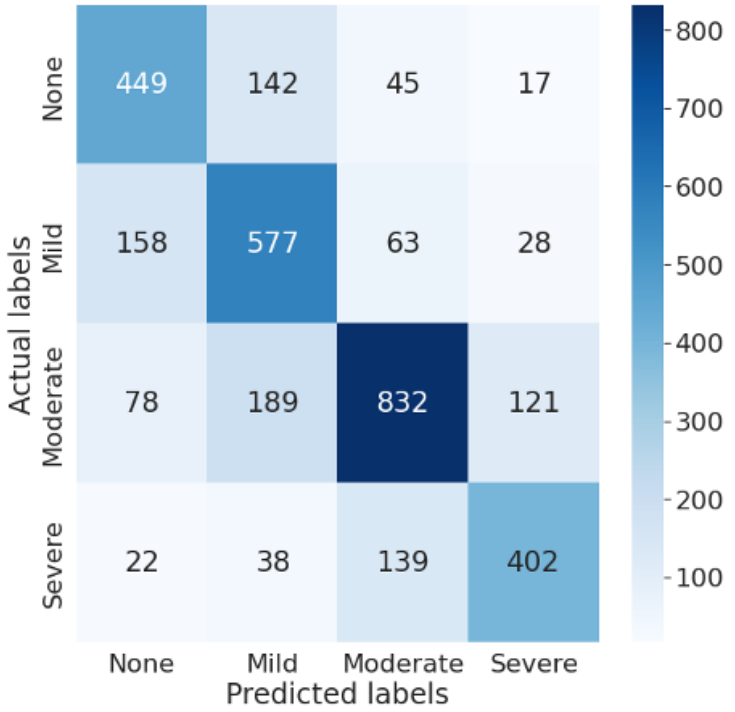
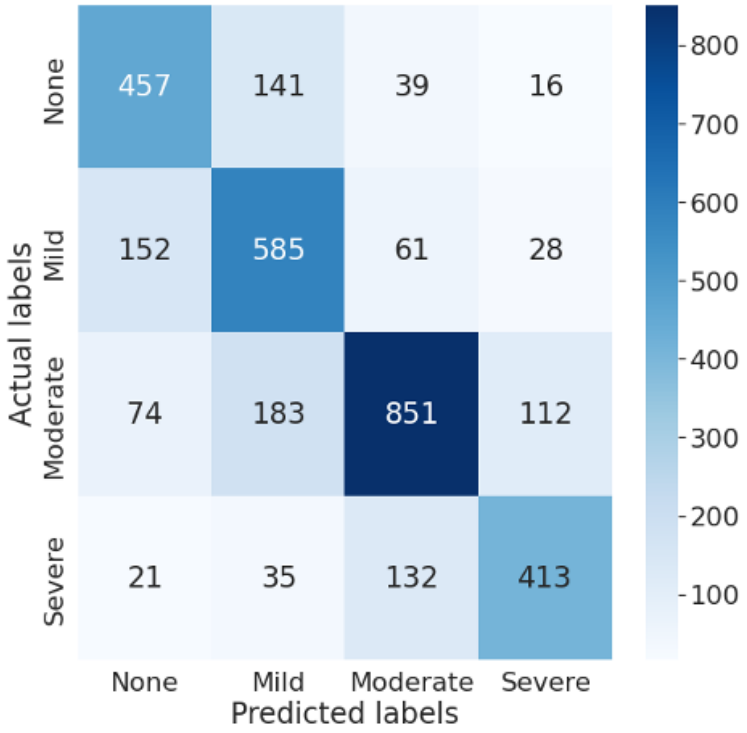


Figure 7. Confusion matrix for the best-performing model with variational autoencoder data representations for 4 pain score levels (pain scores: none=0, mild=1-3, moderate=4-6, and severe=7-10; Table 2, model d).



Role of Medicinal Data

Although prior studies have shown the efficacy of data mining techniques in implementing medical decision-making with treatment outcome prediction [36,37], to the best of our knowledge, this is the first study to analyze the role of medication in pain level prediction for patients with SCD. Our results show that for abstract pain levels, representational learning–based approaches can predict whether a patient is experiencing pain for 22.5% more patients when provided with their medication information (Table 2). This means that when our model is provided with not only vital signs information but also medication type, total medication dosage, and whether the medication was given or missed, it can better predict whether more patients are experiencing pain than when provided with only vital signs. This finding is substantiated by the current medical literature on pain management [38], where clinical research focuses on determining the optimal medication dosage for individual patients. By building a model that incorporates medication information and physiological data, we are one step closer to future pain forecasting that can use current physiological information and pain medication to predict pain at a future time point for assessing the next medication dosage and time.

Furthermore, for a higher resolution of pain levels (ie, 11 levels), our deep representational learning–based approaches could predict subjective pain scores for 13.3% more patients when provided with medication information. In addition, our model can predict such highly subjective pain scores for 38.5% more patients than the random assignment of 9.09% (ie, 1/11 pain scores) when provided with both vital signs and medication information of the patients.

Visualization

We visually inspected the learned representations of the entire data set obtained from VAE representations. Using t-SNE plots, as shown in Figure 3, we compared the disentanglement levels of the represented and raw data. The t-SNE projections clearly show that the VAE can produce sparser and more disentangled representations than the raw data. Although the t-SNE projections of the raw data also indicate data separability, the deep representations can identify variations in mean pain scores (low, moderate, and high). This may explain the competitive performance of the benchmark classifiers in the previous section and the advantage of integrating vital signs and medication data. Although some embeddings were clearly clustered closer to the same pain range, we also observed some overlaps. Specifically, we observed better alignment among the low pain and high pain profiles than among the moderate pain profiles. This may be because of the variation and frequency of the data recordings made for the patients. These preliminary visualization results indicate that our VAE method may require additional data to generate representations that obtain a more granular separation between patients' pain scores.

Patient Profiles

To understand the alignment of the representations learned by our best-performing VAE model, we illustrated 6 sample patient profiles clustered into the 3 pain range categories (no or low=0-3, moderate=4-6, and high=7-10) by the t-SNE projections of the embeddings (as shown in Figure 3). As shown in Table 4, we present 2 patient profiles from each of the 3 categories of pain scores with regard to the medication administered and the vital signs. It should be noted that we specifically chose regions where the pain profiles belonged to 1 of the 3 pain levels. Although we chose 2 patient profiles from

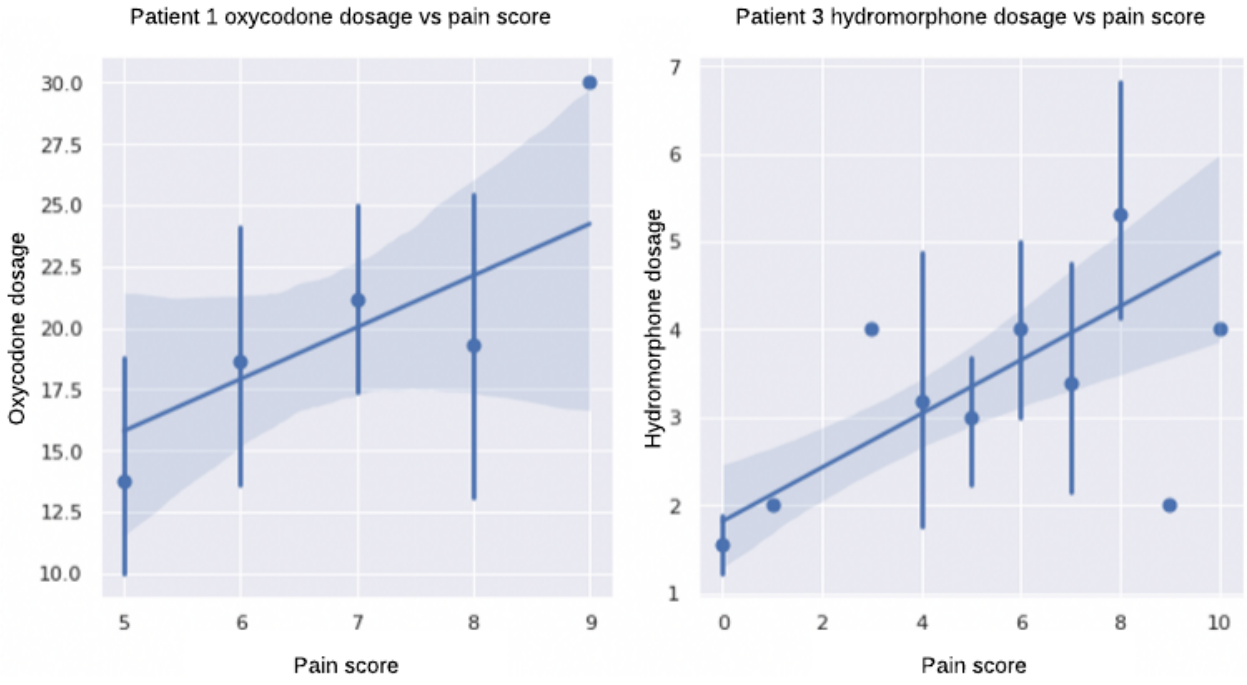
better-aligned regions 1 and 4, we compared 2 patient profiles from a more spread-out moderate pain intensity (regions 2 and 3). Patient numbers are anonymized patient identifiers used in this study.

We observed a positive correlation between the medications administered and pain scores in all 4 patients with high and moderate pain levels. This reflects that the patients reporting higher pain scores were administered an increased medication dosage (as shown in Figure 8 for patients 1 and 3), and our model learned that relationship. For both patients (patients 1 and 2) with high pain, we observed a positive correlation between hydromorphone dosage and pain score, as well as a high correlation between total medication dosage and vital signs, which may be reflective of more pain medications being given when a patient has high pain. This indicates that our model learned the interplay among medications, vital signs, and pain intensity. It will be interesting to analyze these correlations before and after medication in the future.

For both patients (patients 3 and 4) with moderate pain, in addition to a positive correlation between hydromorphone and pain score, we observed a positive correlation between medication and blood pressure. This indicates that our model learned a possible association of administering hydromorphone for moderate pain intensity levels, during which the patients have elevated blood pressure. However, patient 4 had a higher positive correlation between medication dosage and vital signs than patient 3. This may be a possible reason that they were not close in the embedding space and belonged to distant regions, as shown in Figure 3.

Although we did not observe any significant correlation between medication and pain scores for both patients (patients 5 and 6) with no or low pain, we observed a positive correlation between medication and vital signs. This may again be suggestive of elevated vital signs that occur with pain, leading to medication administration. Although both patients might have reported varying pain scores between 0 and 3, it is highly challenging to differentiate between pain scores of 1 and 2 or 2 and 3. Hence, it might be the case that with medication, their vitals improved (as indicated by the positive correlation), making them feel better. This sample patient profile study indicated that deep feature representations can be used to learn complex relationships between various factors influencing pain management. With more data for each patient, this study can be extended to the design of personalized pain management tools to assist clinicians.

Figure 8. Distribution of medication dosage with pain score for sample patients with high and moderate mean pain intensity.



Study Strengths

The design of an objective pain prediction model could potentially assist medical providers in pain management. The lack of objective pain markers has limited the optimal pain assessment strategies for patients with regular pain episodes. As discussed earlier, objective vital signs data can improve pain assessment using ML algorithms. In this study, we developed ML models that can classify pain scores of patients at varying scales and may soon be used to predict pain intensity in individuals with pain based on objective and physiological data and the type, dosage, and status of medication. In the future, a tool designed using our model could be used reasonably quickly to generate pain intensity predictions for unseen new patient data in both inpatient and outpatient hospital visits. This research provides an essential step toward assisting medical practitioners with additional objective pain measures while deciding on a pain management strategy.

Limitations

There were a few limitations to our study. We did not consider our hypothesis that each patient had pain management strategies

through individualized pain protocols, which varied among patients and led to specific pain medications being administered at varying intervals. In addition, because of variations in patient pain record intervals, we did not evaluate before and after administration of pain medication in pain prediction. Both opioid and nonopioid medications, when administered, are known to affect vital sign parameters independently and to varying degrees. For example, opioids can slow a patient's breathing and lower blood pressure. Furthermore, the status of medicine prescription and total medication dosage are subjective variables that may vary between centers and physicians. Our analysis using these variables has not been validated using data from multiple centers.

Furthermore, pain medications may affect patients to different degrees based on the dosing, type of medication, and previous patient history of receiving pain medications. Owing to the variation in data per patient, we could not evaluate such individualized factors. In the future, it will be helpful to analyze the role of individual medication protocols in individualized pain prediction and pre- and postadministration changes.

Conclusions

In this study, we propose an effective pain prediction model based on objective vital signs and pain medication use. Our experiments demonstrated that information about pain medication (type, dosage, and status) can improve pain intensity prediction at both abstract and granular levels. Our analysis indicates the role of medication information in pain assessment and demonstrates that a larger cohort of patient data with deep representational learning improves model performance and can help isolate different patient profiles for further understanding of the role of physiology and medication on pain response. In the future, this study can be extended to further investigate the effect of variation in medication protocols, such as changes in vital signs before and after medication and the time elapsed between medication doses. This would be an essential part of a real-time pain forecasting system and can be extended as a trial that evaluates the timing of the administration of additional doses of opioids based on physiological and objective data alone. Our initial results indicate promise in pursuing each of these efforts, and our study is a valuable addition to ongoing studies investigating how objective vital signs and medication data can be used to help providers to better understand and design pain management strategies.

Acknowledgments

This study was supported by the National Institutes of Health award R01AT010413. Any opinions, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the National Institutes of Health. The authors would like to thank Arvind Subramaniam for data retrieval and Amanuel Alambo for his feedback on this work.

Conflicts of Interest

NS is a speaker and consultant at Novartis and a speaker at Alexion.

References

1. What is sickle cell disease? National Heart, Lung, and Blood Institute. URL: https://www.nhlbi.nih.gov/health/sickle-cell-disease [accessed 2022-04-26]
2. Complications of Sickle Cell Disease and current management approaches. In: National Academies of Sciences, Engineering, and Medicine. Washington, DC: National Academies Press; 2020.
3. Johnson A, Yang F, Gollarahalli S, Banerjee T, Abrams D, Jonassaint J, et al. Use of mobile health apps and wearable technology to assess changes and predict pain during treatment of acute pain in sickle cell disease: feasibility study. JMIR Mhealth Uhealth 2019 Dec 02;7(12):e13671 [FREE Full text] [doi: 10.2196/13671] [Medline: 31789599]
4. Yang F, Banerjee T, Narine K, Shah N. Improving pain management in patients with sickle cell disease from physiological measures using machine learning techniques. Smart Health (Amst) 2018 Jun;7-8:48-59 [FREE Full text] [doi: 10.1016/j.smhl.2018.01.002] [Medline: 30906841]
5. Suzuki T. Does the combination use of two pain assessment tools have a synergistic effect? J Intensive Care 2017;5:1 [FREE Full text] [doi: 10.1186/s40560-016-0195-7] [Medline: 28066556]
6. Padhee S, Alambo A, Banerjee T, Subramaniam A, Abrams D, Nave GJ, et al. Pain intensity assessment in sickle cell disease patients using vital signs during hospital visits. In: Pattern Recognition. Cham: Springer; 2021.
7. Lötsch J, Ultsch A. Machine learning in pain research. Pain 2018 Apr;159(4):623-630 [FREE Full text] [doi: 10.1097/j.pain.0000000000001118] [Medline: 29194126]
8. Alambo A, Andrew R, Gollarahalli S, Vaughn J, Banerjee T, Thirunarayan K, et al. Measuring pain in sickle cell disease using clinical text. In: Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). 2020 Presented at: 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); Jul 20-24, 2020; Montreal, QC, Canada. [doi: 10.1109/embc44109.2020.9175599]
9. Elander J, Midence K. A review of evidence about factors affecting quality of pain management in sickle cell disease. Clin J Pain 1996 Sep;12(3):180-193. [doi: 10.1097/00002508-199609000-00006] [Medline: 8866159]
10. Masi C, Woods K. Variability in inpatient management of sickle-cell disease pain crisis. Clin Res 1993;41(2):A544.

11. Milton JN, Gordeuk VR, Taylor JG, Gladwin MT, Steinberg MH, Sebastiani P. Prediction of fetal hemoglobin in sickle cell anemia using an ensemble of genetic risk prediction models. Circ Cardiovasc Genet 2014 Apr;7(2):110-115 [FREE Full text] [doi: 10.1161/CIRCGENETICS.113.000387] [Medline: 24585758]
12. Patel A, Gan K, Li AA, Weiss J, Nouraie M, Tayur S, et al. Machine-learning algorithms for predicting hospital re-admissions in sickle cell disease. Br J Haematol 2021 Jan;192(1):158-170. [doi: 10.1111/bjh.17107] [Medline: 33169861]
13. Khalaf M, Hussain A, Al-Jumeily D, Keenan R, Keight R, Fergus P, et al. Applied difference techniques of machine learning algorithm and web-based management system for sickle cell disease. In: Proceedings of the International Conference on Developments of E-Systems Engineering (DeSE). 2015 Presented at: International Conference on Developments of E-Systems Engineering (DeSE); Dec 13-14, 2015; Dubai, United Arab Emirates. [doi: 10.1109/dese.2015.18]
14. Abd D, Al-Mejibli I. Monitoring system for sickle cell disease patients by using supervised machine learning. In: Proceedings of the Second Al-Sadiq International Conference on Multidisciplinary in IT and Communication Science and Applications. 2017 Presented at: Second Al-Sadiq International Conference on Multidisciplinary in IT and Communication Science and Applications; Dec 30-31, 2017; Baghdad, Iraq. [doi: 10.1109/aic-mitcsa.2017.8723006]
15. Lazakidou A. Web-based applications in healthcare. In: Web-Based Applications in Healthcare and Biomedicine. Boston, MA, USA: Springer; 2009.
16. Du E, Diez-Silva M, Kato GJ, Dao M, Suresh S. Kinetics of sickle cell biorheology and implications for painful vasoocclusive crisis. Proc Natl Acad Sci U S A 2015 Feb 03;112(5):1422-1427 [FREE Full text] [doi: 10.1073/pnas.1424111112] [Medline: 25605910]
17. Knowlton SM, Sencan I, Aytar Y, Khoory J, Heeney MM, Ghiran IC, et al. Sickle cell detection using a smartphone. Sci Rep 2015 Oct 22;5:15022 [FREE Full text] [doi: 10.1038/srep15022] [Medline: 26492382]
18. Tighe PJ, Harle CA, Hurley RW, Aytug H, Boezaart AP, Fillingim RB. Teaching a machine to feel postoperative pain: combining high-dimensional clinical data with machine learning algorithms to forecast acute postoperative pain. Pain Med 2015 Jul;16(7):1386-1401 [FREE Full text] [doi: 10.1111/pme.12713] [Medline: 26031220]
19. Arbour C, Choinière M, Topolovec-Vranic J, Loiselle CG, Gélinas C. Can fluctuations in vital signs be used for pain assessment in critically ill patients with a traumatic brain injury? Pain Res Treat 2014;2014:175794 [FREE Full text] [doi: 10.1155/2014/175794] [Medline: 24639895]
20. Acute Pain Management: Scientific Evidence Fifth edition 2020. Melbourne, Australia: Australian and New Zealand College of Anaesthetists and Faculty of Pain Medicine; 2020.
21. Nickerson P, Tighe P, Shickel B, Rashidi P. Deep neural network architectures for forecasting analgesic response. Annu Int Conf IEEE Eng Med Biol Soc 2016 Aug;2016:2966-2969 [FREE Full text] [doi: 10.1109/EMBC.2016.7591352] [Medline: 28268935]
22. Payne J, Aban I, Hilliard LM, Madison J, Bemrich-Stolz C, Howard TH, et al. Impact of early analgesia on hospitalization outcomes for sickle cell pain crisis. Pediatr Blood Cancer 2018 Dec;65(12):e27420 [FREE Full text] [doi: 10.1002/pbc.27420] [Medline: 30151977]
23. Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (ehr) analysis. IEEE J Biomed Health Inform 2018 Sep;22(5):1589-1604 [FREE Full text] [doi: 10.1109/JBHI.2017.2767063] [Medline: 29989977]
24. Miotto R, Li L, Kidd BA, Dudley JT. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. Sci Rep 2016 May 17;6:26094 [FREE Full text] [doi: 10.1038/srep26094] [Medline: 27185194]
25. Choi E, Bahadori MT, Schuetz A, Stewart WF, Sun J. Doctor AI: predicting clinical events via recurrent neural networks. JMLR Workshop Conf Proc 2016 Aug;56:301-318 [FREE Full text] [Medline: 28286600]
26. Kingma D, Welling M. Auto-encoding variational Bayes. arXiv. 2013 Dec 20. URL: http://arxiv.org/abs/1312.6114v10 [accessed 2022-01-02]
27. Rezende D, Mohamed S, Wierstra D. Stochastic Backpropagation and Approximate Inference in Deep Generative Models. In: Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32. 2014 Presented at: 31st International Conference on International Conference on Machine Learning - Volume 32; Jun 21–26, 2014; Beijing, China.
28. Mattei PA, Frellsen J. MIWAE: deep generative modelling and imputation of incomplete data sets. In: Proceedings of the 36th International Conference on Machine Learning. 2019 Presented at: 36th International Conference on Machine Learning; May 24, 2019; Long Beach, California.
29. Nazábal A, Olmos P, Ghahramani Z, Valera I. Handling incomplete heterogeneous data using VAEs. Pattern Recognition 2020 Nov;107:107501 [FREE Full text] [doi: 10.1016/j.patcog.2020.107501]
30. Sadati N, Nezhad M, Chinnam R, Zhu D. Representation learning with autoencoders for electronic health records: a comparative study internet. arXiv. 2018. URL: http://arxiv.org/abs/1801.02961 [accessed 2022-01-04]
31. Kingma DP, Welling M. An Introduction to Variational Autoencoders. FNT in Machine Learning 2019 Jun 6;12(4):307-392 [FREE Full text] [doi: 10.1561/2200000056]
32. Ho T. Random decision forests. In: Proceedings of 3rd International Conference on Document Analysis and Recognition. 1995 Presented at: 3rd International Conference on Document Analysis and Recognition; Aug 14-16, 1995; Montreal, QC, Canada. [doi: 10.1109/icdar.1995.598994]

33. Tibshirani R. Regression shrinkage and selection via the lasso: a retrospective. J Royal Stat Soc Series B (Methodological) 2011 Apr 20;73(3):273-282. [doi: 10.1111/j.1467-9868.2011.00771.x]
34. Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995 Sep;20(3):273-297. [doi: 10.1007/BF00994018]
35. Van der Maaten L, Hinton G. Visualizing data using t-SNE. J Mach Learn Res 2008;9:2579-2605 [FREE Full text]
36. Parrales Bravo F, Del Barrio García AA, Gallego MM, Gago Veiga AB, Ruiz M, Guerrero Peral A, et al. Prediction of patient's response to OnabotulinumtoxinA treatment for migraine. Heliyon 2019 Feb;5(2):e01043 [FREE Full text] [doi: 10.1016/j.heliyon.2018.e01043] [Medline: 30886915]
37. Athreya A, Iyer R, Neavin D, Wang L, Weinshilboum R, Kaddurah-Daouk R, et al. Augmentation of physician assessments with multi-omics enhances predictability of drug response: a case study of major depressive disorder. IEEE Comput Intell Mag 2018 Aug;13(3):20-31 [FREE Full text] [doi: 10.1109/MCI.2018.2840660] [Medline: 30467458]
38. Lakkakula BV, Sahoo R, Verma H, Lakkakula S. Pain management issues as part of the comprehensive care of patients with sickle cell disease. Pain Manag Nurs 2018 Dec;19(6):558-572. [doi: 10.1016/j.pmn.2018.06.004] [Medline: 30076112]

Abbreviations

AUC: area under the curve
EHR: electronic health record
ML: machine learning
RBC: red blood cell
RF: random forest
SCD: sickle cell disease
SD: Standard Deviation
SVM: support vector machine
t-SNE: t-distributed stochastic neighboring embedding
VAE: variational autoencoder

Edited by A Mavragani; submitted 02.02.22; peer-reviewed by S Badawy, H Park, A Staffini; comments to author 06.04.22; revised version received 27.04.22; accepted 08.05.22; published 23.06.22.

Please cite as:
Padhee S, Nave Jr GK, Banerjee T, Abrams DM, Shah N
Improving Pain Assessment Using Vital Signs and Pain Medication for Patients With Sickle Cell Disease: Retrospective Study
JMIR Form Res 2022;6(6):e36998
URL: https://formative.jmir.org/2022/6/e36998
doi:10.2196/36998
PMID:35737453

Original Paper

Predicting Depression in Adolescents Using Mobile and Wearable Sensors: Multimodal Machine Learning–Based Exploratory Study

Tahsin Mullick[1], MEng; Ana Radovic[2], MD, MSc; Sam Shaaban[3], MBA; Afsaneh Doryab[1], PhD

[1]Department of Engineering Systems and Environment, University of Virginia, Charlottesville, VA, United States

[2]Department of Pediatrics, University of Pittsburgh, Pittsburgh, PA, United States

[3]NuRelm, Pittsburgh, PA, United States

Corresponding Author:
Tahsin Mullick, MEng
Department of Engineering Systems and Environment
University of Virginia
Olsson Hall, 151 Engineer's Way
Charlottesville, VA, 22904
United States
Phone: 1 434 243 5823
Email: tum7q@virginia.edu

Abstract

Background: Depression levels in adolescents have trended upward over the past several years. According to a 2020 survey by the National Survey on Drug Use and Health, 4.1 million US adolescents have experienced at least one major depressive episode. This number constitutes approximately 16% of adolescents aged 12 to 17 years. However, only 32.3% of adolescents received some form of specialized or nonspecialized treatment. Identifying worsening symptoms earlier using mobile and wearable sensors may lead to earlier intervention. Most studies on predicting depression using sensor-based data are geared toward the adult population. Very few studies look into predicting depression in adolescents.

Objective: The aim of our work was to study passively sensed data from adolescents with depression and investigate the predictive capabilities of 2 machine learning approaches to predict depression scores and change in depression levels in adolescents. This work also provided an in-depth analysis of sensor features that serve as key indicators of change in depressive symptoms and the effect of variation of data samples on model accuracy levels.

Methods: This study included 55 adolescents with symptoms of depression aged 12 to 17 years. Each participant was passively monitored through smartphone sensors and Fitbit wearable devices for 24 weeks. Passive sensors collected call, conversation, location, and heart rate information daily. Following data preprocessing, 67% (37/55) of the participants in the aggregated data set were analyzed. Weekly Patient Health Questionnaire-9 surveys answered by participants served as the ground truth. We applied regression-based approaches to predict the Patient Health Questionnaire-9 depression score and change in depression severity. These approaches were consolidated using universal and personalized modeling strategies. The universal strategies consisted of Leave One Participant Out and Leave Week X Out. The personalized strategy models were based on Accumulated Weeks and Leave One Week One User Instance Out. Linear and nonlinear machine learning algorithms were trained to model the data.

Results: We observed that personalized approaches performed better on adolescent depression prediction compared with universal approaches. The best models were able to predict depression score and weekly change in depression level with root mean squared errors of 2.83 and 3.21, respectively, following the Accumulated Weeks personalized modeling strategy. Our feature importance investigation showed that the contribution of screen-, call-, and location-based features influenced optimal models and were predictive of adolescent depression.

Conclusions: This study provides insight into the feasibility of using passively sensed data for predicting adolescent depression. We demonstrated prediction capabilities in terms of depression score and change in depression level. The prediction results revealed that personalized models performed better on adolescents than universal approaches. Feature importance provided a better understanding of depression and sensor data. Our findings can help in the development of advanced adolescent depression predictions.

(JMIR Form Res 2022;6(6):e35807) doi:10.2196/35807

KEYWORDS

adolescent; depression; uHealth; machine learning; mobile phone

Introduction

Background

According to the World Health Organization, half of all mental health conditions start at the age of 14 years, but most cases are undetected and untreated. Among mental health conditions, depression is one of the leading causes of illness and disability among adolescents [1], the most likely mental illness to be a risk factor for suicide [2], the second leading cause of death among US adolescents [3], and among the top causes of death in adolescents worldwide [4].

Major depressive disorder, more commonly termed *depression*, can be defined as a medical disorder that results in negative feelings in a person's thoughts or actions. The effects of depression are both emotional and physical [5]. The sources of depression are varied and include biochemical changes, genetics, personality traits, and environmental factors [5]. Depression has a combination of effects that play a role in its diagnosis, such as alteration in mood, negative self-image, self-punitive desires, vegetative changes, and physiological changes such as activity retardation or agitation [6].

Depression is difficult to monitor or regulate in adolescents as part of the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, diagnosis [7] includes not only depressive symptoms but also irritability, which may be difficult to distinguish from typical adolescent behavior. As an internalizing disorder that is expressed more through thoughts and not actions, worsening depressive symptoms can be more difficult for others such as parents or caregivers to identify [8]. Adolescents also report using cognitive coping strategies far less often than adults [9]. In a study on adolescent mental health literacy, it was found that <50% of adolescents were able to identify depression [10]. Although earlier intervention on symptom worsening improves outcomes in depression, adolescents themselves and their caregivers not being able to identify these symptoms serves as a barrier [11]. This indicates a strong need for interventions that can assist adolescents and their caregivers in monitoring the symptoms of depression earlier.

The result of not addressing depression can extend into adulthood, impairing both physical and mental health and limiting future employment opportunities and the potential to lead satisfied lives [12]. With the increased use of screening tools such as the Patient Health Questionnaire-9 (PHQ-9), mental health clinicians and primary care providers can more efficiently screen for depression. However, screening does not always lead to a substantial increase in treatment engagement [13]. Measurement-based care [14] or using these validated screening tools as recurring to identify, monitor, and treat depressive symptoms results in improved outcomes for patients with depression by identifying and intervening earlier on worsening nonresponsive symptoms or their treatment [15].

The success of validated screening tools has provided mental health clinicians and primary care providers with better assessment tools for symptom severity and, especially with the ability to embed these tools in electronic health records, more frequent monitoring may result in earlier intervention and improved care. Unfortunately, constant monitoring of depression symptomatology is still far from reality. With the advent of mobile phones, fitness trackers, and their inbuilt sensors, this can be made possible. Our goal is to look closely into adolescent depression through the eyes of passively sensed data and evaluate machine learning (ML) approaches that offer predictive capabilities.

By exploring approaches to adolescent depression data, we want to enable the future development of apps geared toward the continuous monitoring of patients experiencing depression and allow clinicians, adolescents, and their parents the opportunity to take preventive or earlier actions.

This study was aimed at using passively sensed data to generate predictions on depression levels and change in depression levels. The predictions took on both universal and personalized modeling approaches. We then determined key contextual features that affected our ML models. Finally, we presented how the performance of personalized models changed over time and across data samples.

Related Work

Related work in this section takes an inverse pyramid approach to describe the state of the art in mobile sensing for health apps and then focuses on the impact in the space of mental health.

Mobile-Based Sensing for Health Apps

Mobile sensing has been an active research area in health apps. A number of studies [16-18] have analyzed areas of cardiovascular health and sensed participant heart rate and heart rate variability with the help of mobile camera sensors. Areas of study such as sleep have benefited from mobile sensing by using sensors to detect sleep quality and sleep states using ML [19]. Further studies on sleep have explored both supervised and unsupervised approaches to detect sleep variation in contextual settings [20-22]**.** Mobile sensors such as accelerometers, gyroscopes, and GPSs have been used to model human behavior and cognition through contextualized feature extraction [23,24]. Studies on overall health and well-being have combined the aforementioned sensing capabilities to help promote general health. For example, the use of health apps to monitor human behavior through sleep, physical activity, and social interaction [25] has been found to show improvement in behavior patterns. Another example of general well-being [26] generates an index as a medium of feedback for improving health through exercise-based goal setting. All of the aforementioned studies have shown the efficacy of using mobile sensing to predict or diagnose health-related changes. The next subsection delves into how mobile sensing is changing mental health.

Mobile Sensing in Mental Health Apps

Mobile sensing–based mental health studies have been conducted in the areas of bipolar disorder [27], schizophrenia [28], anxiety [29,30], stress [31,32], and depression [33-39]. These studies have shown that mobile sensing can play an integral role in detecting and predicting mental health–related problems. Daily mood, physical activity, and social communication tracking of participants helped predict symptoms of bipolar relapse [27]. This was achieved using random coefficient methods to analyze the relationship between phone-based data and the rating of manic and depressive symptoms. Schizophrenia is another mental condition in which passive sensing has demonstrated predictive capability by showing the relationships between tracked features as indicators of schizophrenia [28]. The study used bivariate analysis and tree-based methods to perform ecological momentary assessment scores. Depression and anxiety in college students were studied, in particular the effect of stress and self-esteem, using the tool of causal networks derived from time-series sensor data [29]. These data helped in understanding the causal relationship between anxiety, depression, and stress. Anxiety regulation using wearable devices was explored through false feedback of slow heart rate [30] and was found to be beneficial for helping control anxiety symptoms. Researchers have been successful in tracking physiological changes during stress using voice sensing across different acoustic environments and individuals [31]. Patients undergoing chemotherapy were studied using passively sensed data. This exploratory study used instruments of random forest classifiers showing a strong correlation between sedentary behavior, less time spent in light physical activity, and other factors such as longer onscreen time and app interactions [32]. All of these studies provide sufficient evidence to consider passively sensed data as an effective method to track mental health, which provides support for our approach in this study.

One of the first studies to use mobile phones for depression used GPS data to track participant mobility [33]. The study provided evidence of a correlation between location-based data and depressive mood. In addition to GPS, phone use has been another feature to exhibit a strong relationship with depression severity [34]**.** The study extracted features such as phone use frequency and duration along with GPS-based features such as location variance and normalized entropy to show the correlation with depression. Behavior in people with depression has also been investigated by monitoring additional features such as sleep and social interaction through smartphones [35]. Multimodal features were gradually introduced into the research sphere to derive a contextual filtering of features that detected depression in college students and showed that multimodal feature information could outperform unimodal features [36]. This study used association rule mining to choose features and applied standard ML to detect depression, showing the merit in using multimodal features. Detecting depression is dependent on the approach used; analyzing the problem from the perspective of longitudinal data and exploring changes in depressive symptoms were shown to generate good accuracy [37]. The work in the latter study is closely related to our endeavor and serves as an inspiration. Collaborative filtering-based study is yet another approach that has shown promise in using personalized models to derive better predictions [38]. Our study also proposes 2 personalized strategies to model individual participants using ML.

Narrowing down to adolescent depression studies, we present some existing works in the literature and later explain their differences from our work in Table 1. Studies on adolescent depression have been primarily survey-based and social sentiment– and feasibility-centric [40-42]. The work in the study by Cao et al [39] closely relates to our aim of detecting depression in adolescents. However, their study had a smaller sample size and used both parent and adolescent inputs and was also more reliant on participant feedback. The differences between the studies highlighted and our work are further elaborated on in the *Discussion* section.

In this study, we first investigated the feasibility of universal and personalized ML modeling strategies to predict adolescent depression scores and change in depression levels. We then identified features that were more predictive of adolescents' depression during the ML process. Finally, we studied how missingness of data affected model performance along with understanding how much data were required for our models to perform over a predetermined threshold.

Our findings revealed that a regression-based predictive modeling approach was able to capture more granular changes in depression scores. We also showed that personalized strategies were more effective predictors compared with universal strategies. The performance of personalized models did not improve with increase in the weeks of data and, instead of a steady increase in model performance with increase in data, we experienced fluctuations in the results. In an attempt to explain this phenomenon, we performed additional studies that separated our participants into 2 pools. The pools were generated based on the SD of the depression scores. Our results showed that the pool with a small SD in depression score was more accurately modeled in comparison with the higher-SD pool.

Table 1. Papers on adolescent mental health prediction and how our work differs from the existing work.

Paper	Study aim	Methods	Results	Difference from our work
Cao et al [39]	Investigated the effectiveness of smartphone apps useful in evaluating and monitoring depression symptoms in a clinically depressed adolescent population compared with psychometric instruments (PHQ-9[a], HAM-D[b], and HAM-A[c]); 13 participants aged 12 to 17 years	Used self-evaluation of adolescents and parents with smartphone data to improve predictions of PHQ-9 scores; used the SOLVD app installed only on Android phones; used only linear regressor and support vector regressor with polynomial kernel	Correlation between mood averaged over a 2-week period and biweekly psychometric score from PHQ-9, HAM-D, and HAM-A; combining self-evaluation from both parents and children along with smartphone sensor data resulted in PHQ-9 score prediction accuracy	Our work does not depend on self-evaluation by adolescents and parents to help improve predictions; instead, we consider a system where our reliance is exclusively on the captured sensor values to make predictions of PHQ-9 scores. We used universal and personalized modeling strategies with multiple machine learning algorithms.
Maharjan et al [43]	StandStrong app used to assess feasibility and acceptability of sensing technologies for maternal depression treatment in low-resource settings for mothers aged between 15 and 25 years	They explored possible explanations for differences in successful data collection by time of day and sensor type along with description of qualitative results to illuminate these differences	The study mainly identified concerns related to technological barriers in passively sensed data collection.	The study was based on passively sensed data collection. It did not perform predictive modeling. The aim was to assess how well the app performed in data collection and the hurdles encountered therein. The study had 11 participants with depression with a mix of young and older participants, whereas our study was focused on adolescents, and all participants had been diagnosed with some form of depression.
MacLeod et al [44]	Explored whether passively collected smartphone sensor data can be used to predict internalizing symptoms among youths in Canada; participants aged between 10 and 21 years	Self-reports of anxiety, depression, and attention-deficit hyperactivity disorder collected; N=122 for 2 weeks of passively sensed data; CES-DC[d] and SCARED[e] anxiety assessments were used	Depressive symptoms correlated with time spent stationary, less mobility, higher light intensity during the night, and fewer outgoing calls. Anxiety correlated with less time spent stationary, greater mobility, and more time on-screen. Adding passively collected smartphone data to prediction models of internalizing symptoms significantly improved their fit.	This work was primarily focused on establishing correlations between self-reports. The study used passive sensor data to perform linear regressor model fitting for predictions of the CES-DC and SCARED values. Nonlinear modeling approaches were not considered, whereas we have explored and produced better results.

[a]PHQ-9: Patient Health Questionnaire-9.

[b]HAM-D: Hamilton Depression Rating Scale.

[c]HAM-A: Hamilton Anxiety Rating Scale.

[d]CES-DC: Center for Epidemiological Studies Depression Scale for Children.

[e]SCARED: Screen for Child Anxiety Related Disorders.

Methods

Data Collection

Recruitment and Participant Breakdown

Adolescents aged 12 to 17.99 years and their parents were recruited from psychiatric clinics at the University of Pittsburgh Medical Center Western Psychiatric Hospital serving depressed and suicidal youth, an adolescent and young adult medicine clinic seeing youth for primary and subspecialty services, as well as through the University of Pittsburgh research registry. A total of 114 adolescents expressed an interest in this study. Of these 114 adolescents, 94 (82.5%) completed a screening assessment, and 31 (27.2%) were screened out because of minimal symptoms of depression (PHQ-9 score [45] of <5), no self-reported previous diagnosis of depression, not having a smartphone, and age restrictions. A total of 57 adolescents and their parents consented to the study, of whom 55 (96%) completed a baseline assessment and were entered into the study. The aggregated data set after exploratory data analysis (EDA) and initial cleaning consisted of 67% (37/55) of the participants. The reduction in participant number was due to sensor issues and irregular syncing that constituted missing data and the dropping out of some participants in between the study. The data for each participant were collected over a period of 24 weeks.

Passively sensed data from mobile phones were collected using the AWARE app [24], which logs relevant sensor data and harnesses those data within the device. It was installed on the participants' phones and set up to record the sensor information

in the desired sampling frequencies. We collected data from multiple sensors, including calls, conversations, location, Wi-Fi, and screen use. Features were classified into event-based features, which included phone use, calling, and conversational recording, and time series–based features, comprising Wi-Fi and GPS-based location. We used Fitbit Inspire HR (software version 1.84.5) to collect heart rate, sleep, and steps. Sensor data from GPS and Wi-Fi were collected at a frequency of 10 minutes. The Fitbit features were collected every minute and accumulated daily. The data collected from both AWARE and Fitbit were uploaded to the cloud and then hosted in a database for cleaning and further processing.

The AWARE passive sensing data and Fitbit were, on average, 69.11% and 32.36% complete, respectively. Missing Fitbit data were attributed to less than expected adherence to wearing the Fitbit because of several reasons, including forgetting to wear it, fatigue, rash (recurred in 1/55, 2% of the participants even after the band was changed), and the need to charge the device. The data collection process was approved by the University of Pittsburgh Institutional Review Board.

Weekly PHQ-9 surveys were sent over the 24 weeks, and the adolescents completed 69.01% (873/1265) of the weekly surveys on average, respectively. The PHQ-9 is an evaluative questionnaire used to assess depression severity. The PHQ-9 has been used effectively in multiple studies related to depression [38,39]. The questionnaire consists of a set of 9 questions with scores between 0 and 3. This results in an overall score range between 0 and 27. For the purpose of our study, this was our choice ground truth owing to its strength in categorizing depression severity levels and its effectiveness in yielding responses from participants when administered remotely [46,47]. The scores are divided into levels based on depression severity and allow for easier interpretability by clinicians, parents, and adolescents [45].

Descriptive Statistics of Collected Participant Data

The adolescent sample included participants aged 12 to 17.99 years, with an average age of 15.5 years. Most of the sample was White (47/56, 84%), with 16% (9/55) of the individuals representing a minority population. There was variability in gender, with approximately 73% (41/56) of the adolescent sample identifying as female, 23% (13/56) identifying as male, and 9% (3/56) identifying as transgender or other. The demographic statistics are provided in Figure 1 associated with the data collected.

Figure 1. Demographic statistics: (A) gender distribution, (B) race distribution of the adolescents, (C) sexual orientation, (D) depression score distribution for each week of observation, and (E) depression score distribution for each participant.

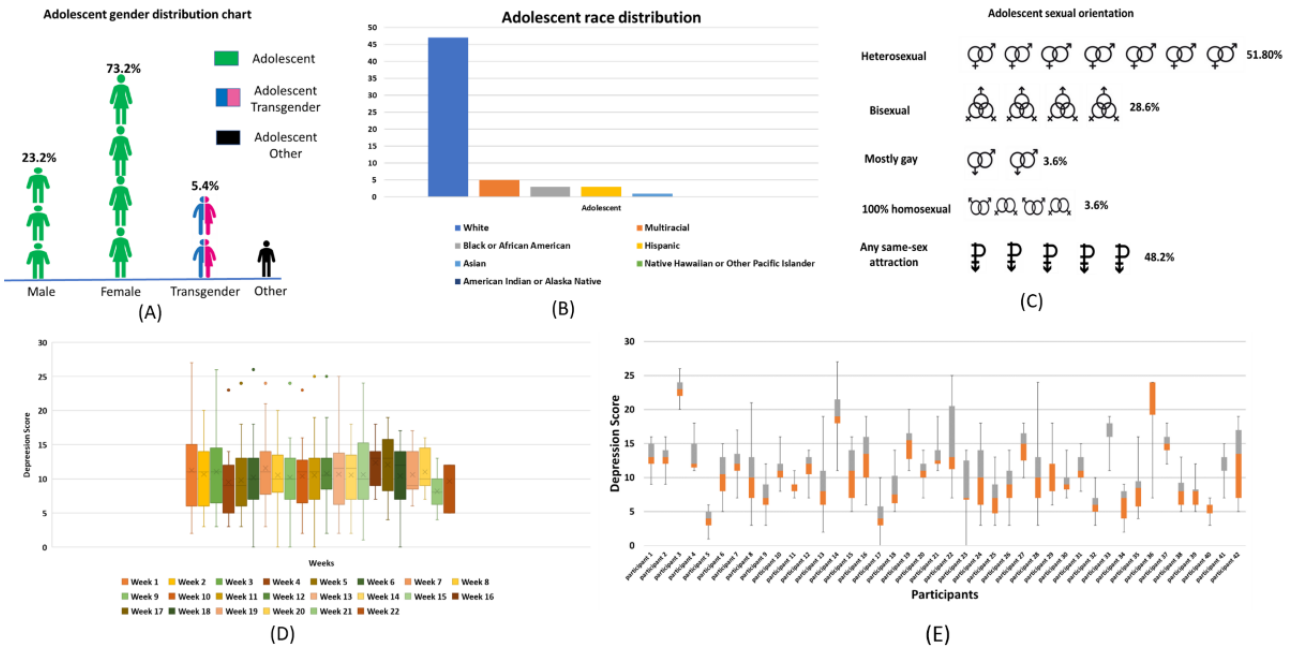


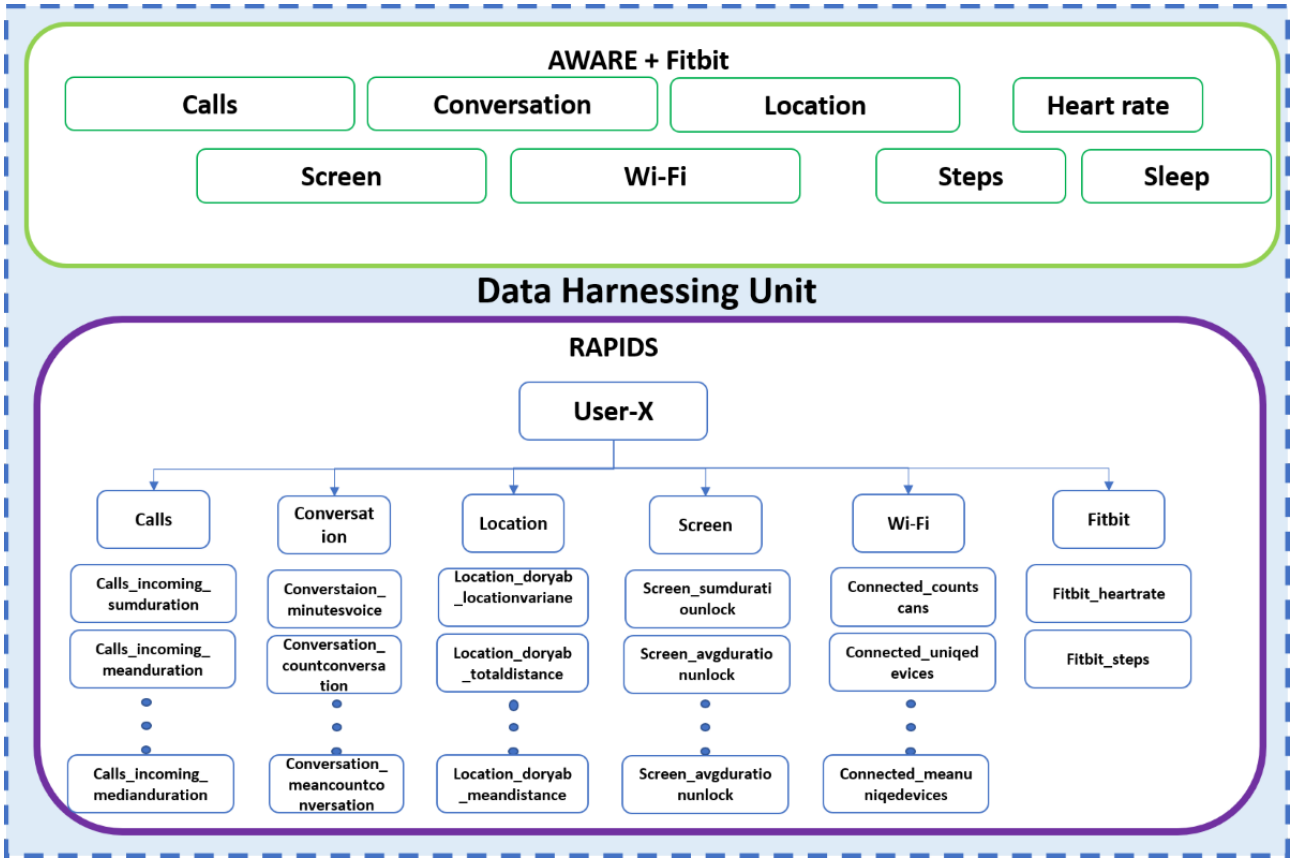
Figure 1 also contains box plots of depression scores based on weeks (bottom left) and depression scores based on the participants' box plots (bottom right). The depression score versus participants box plot presents the variation in depression scores across participants. The data set comprised 507 data points. The PHQ-9 scores ranged from a minimum of 0 to a maximum of 27. The mean PHQ-9 score was 11.21 (SD 5.23). For depression score versus weeks, we observed a mean PHQ-9 score of 10.63 (SD 4.92), and the minimum and maximum values were similar to those of the participant plot. The PHQ-9 scores are also expressed in the form of levels of depression: minimal (0-4), mild (5-9), moderate (10-14), moderately severe (15-19), and severe (20-27). The distribution of depression levels according to the number of participants was as follows: minimal depression (12/55, 22%), mild depression (26/55, 47%), moderate depression (31/55, 56%), moderately severe depression (21/55, 38%), and severe depression (5/55, 9%). There were rare occurrences of participants traversing up to 4 levels of depression over the course of their time in the study. It is also important to mention that, owing to data limitations and survey completion rate, 5% (3/55) of the participants maintained a single level of depression in the data set. The observations also revealed that most participants fluctuated between 2 levels of depression.

Feature Extraction

The collected sensor data were passed to the Reproducible Analysis Pipeline for Data Streams framework [25] for feature extraction. The data set retained 66 features, including calls, conversations, locations, screen, Wi-Fi, heart rate, sleep, and steps. The data were then compiled into an aggregated data set and used as input for our ML modeling operations.

The data set was in a 2D tabular format for the application of our supervised modeling approaches populated with the survey results from the PHQ-9 weekly surveys to serve as the ground truth. To match the weekly ground truth depression score, we aggregated our features into daily and then weekly values. Figure 2 shows the combined harnessing framework comprising AWARE, Fitbit, and Reproducible Analysis Pipeline for Data Streams. Each sensor-based feature set was used to extract a range of features.

Figure 2. Feature extraction.



ML Modeling

Overview

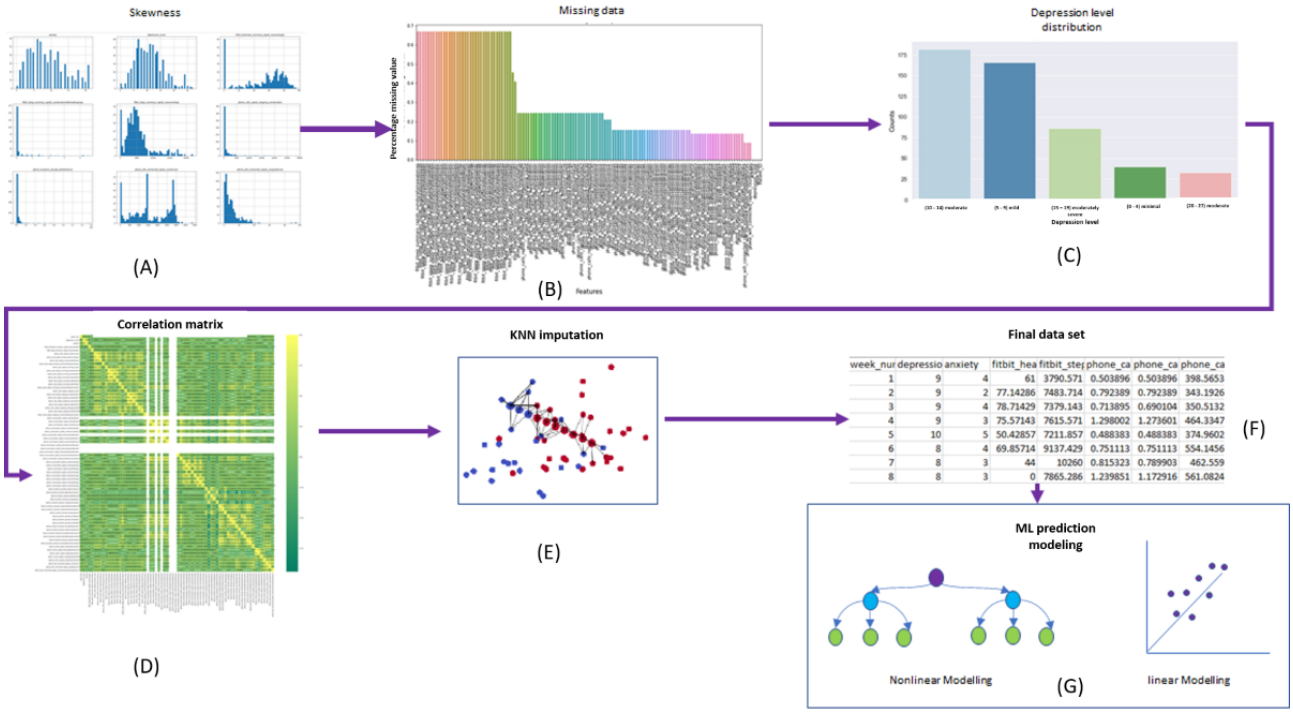
The data processing pipeline started with extensive EDA to check for skewness and filter missing data. This step was followed by the calculation of the Pearson correlation values for our feature set and the removal of highly correlated features. On the basis of our EDA, we set thresholds for missing data and adopted a robust imputation strategy such as the k-nearest neighbors, which is effective in handling multivariate time-series data. Our final data set consisted of 507 data points with 61 features, which represented 37 participants owing to high data sparsity. An illustration of the EDA and final data set generation is presented in Figure 3.

The ML phase after the data preprocessing can be segmented into a model-fitting stage and a cross-validation (CV) stage.

In the model-fitting stage, we applied both the depression score prediction and change in depression level prediction approaches. This model involved passing the feature sets through linear and nonlinear ML algorithms. The linear algorithms included Least Absolute Shrinkage and Selection Operator and elastic net. Nonlinear modeling included tree-based algorithms such as random forest; decision trees; and ensemble methods such as AdaBoost, extra trees, gradient boosting, and XGBoost.

The CV stage was responsible for the train-test splitting of data. This stage was also designed to consider universal and personalized modeling strategies. The universal strategies ensured that the modeling was based on the sample population data splits. The personalized strategies modeled based on individual data train-test splits. These strategies will be further elaborated on in the following subsection.

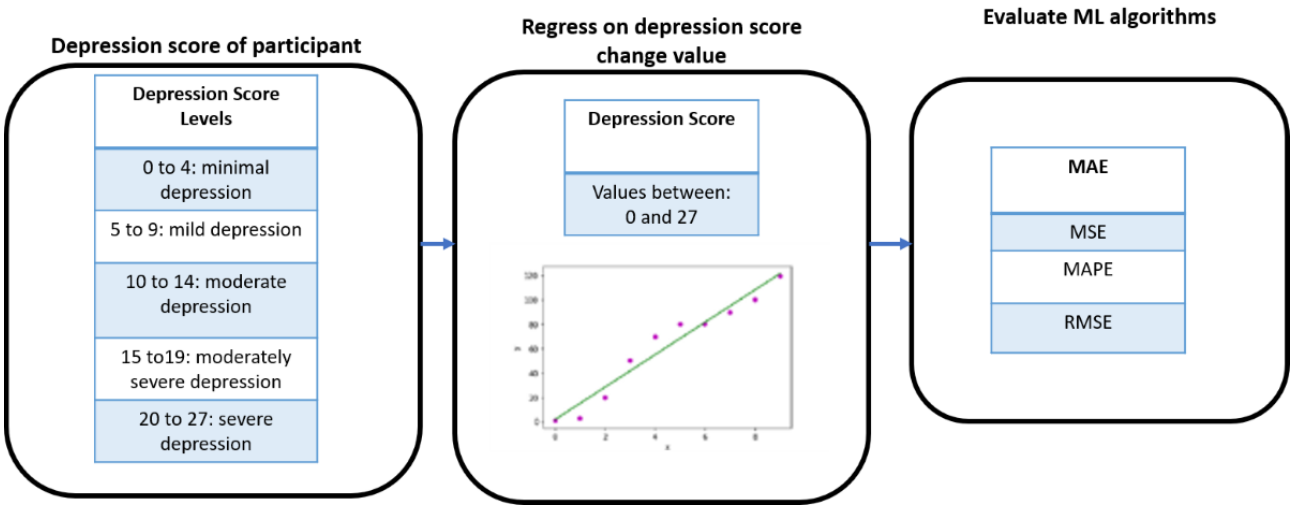
Figure 3. Machine learning (ML) pipeline comprising exploratory data analysis that includes (A) check for skewness of data, (B) missing value assessment, (C) check of depression level distribution, (D) generation of correlation matrix and removal of features that are highly correlated, (E) k-nearest neighbors (KNN)-based missing value imputation, (F) aggregated data set creation, and (G) nonlinear and linear ML modeling of data.



Prediction of Depression Score

To predict depression score, we used linear and nonlinear regression-based ML algorithms, as shown in Figure 4. The algorithms included Least Absolute Shrinkage and Selection Operator, elastic net, random forest, AdaBoost, extra trees, gradient boosting, and XGBoost for regression. The features extracted were used as input based on sensor combinations. The ML algorithms modeled on the data output predictions of the depression score. The model was based on universal and personalized modeling strategies. The models were evaluated based on mean absolute error (MAE), mean squared error (MSE), root MSE (RMSE), and mean absolute percentage error (MAPE).

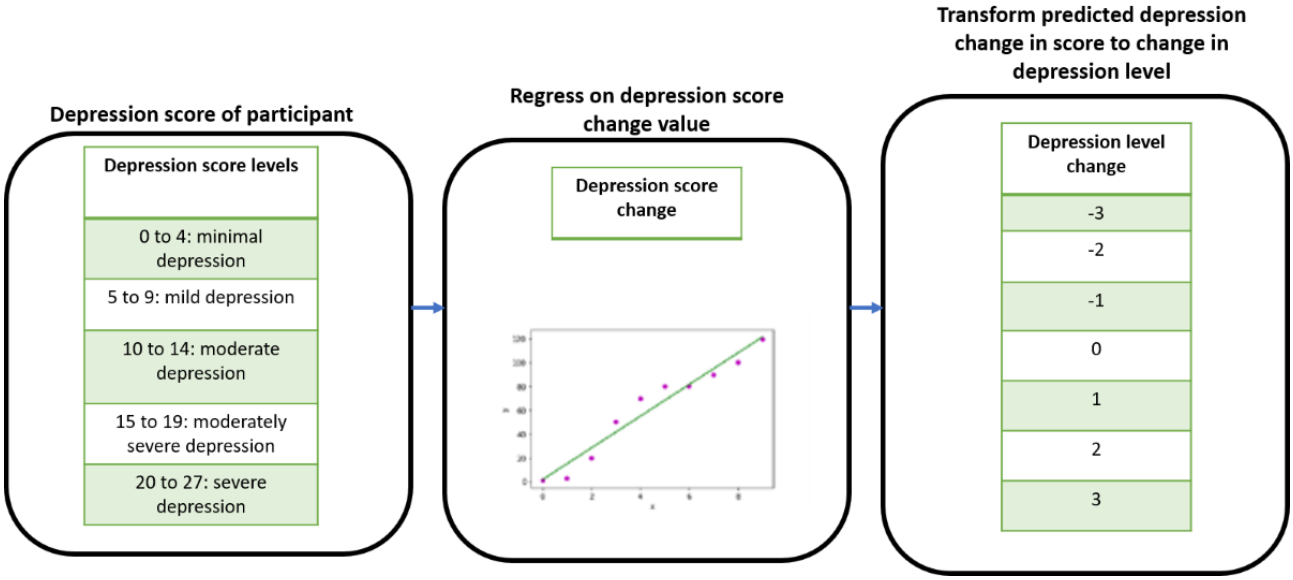
Figure 4. Depression score prediction approach. MAE: mean absolute error; MAPE: mean absolute percentage error; ML: machine learning; MSE: mean squared error; RMSE: root mean squared error.



Prediction of Change in Depression Level

The prediction of change in depression level used the feature set combinations as input. The ML algorithms regressed on the feature data to predict the change in depression score and is shown in Figure 5. The change in depression level was then derived from the predicted change in depression score. This was a regression modeling approach with MAE, MSE, RMSE, and MAPE as evaluation metrics.

Figure 5. Machine learning approach for predicting change in depression level.

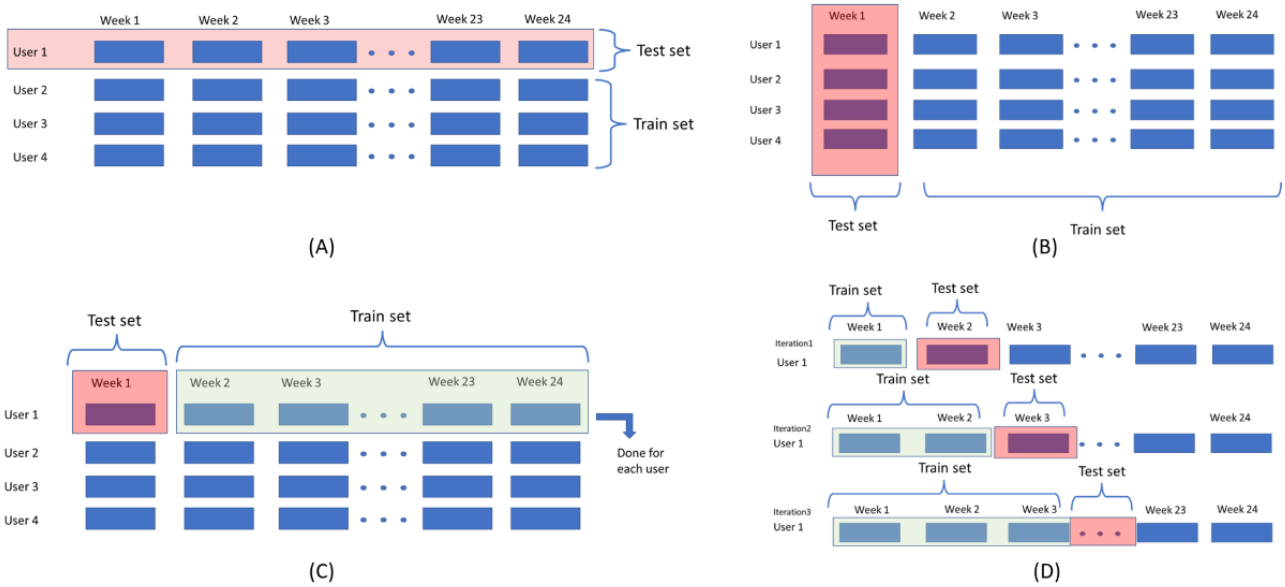


As mentioned previously, there were 5 depression levels. A jump to a level above (positive change) or a level below (negative change) was considered a change in level. The magnitude of the change was determined by the number of jumps seen in participant depression levels. On the basis of this idea, there were 9 changes in levels: positive changes (1, 2, 3, and 4), negative changes (−1, −2, −3, and –4), and no change (0). The change in depression level observed in our data fell within the range of −3 to 3. The change in depression scores was mapped to these 7 changes in depression levels. The establishment of levels helps in the better interpretation of depression changes by health care providers and aligns with standard medical diagnostics [45].

CV Strategy

We used multiple variations of the leave-one-out CV as presented in Figure 6. These strategies were designed to accommodate both personalization and generalization of the trained models.

Figure 6. Cross-validation strategies: (A) Leave One User Out, (B) Leave Week X Out, (C) Leave One Week One User Instance, and (D) Accumulated Weeks.



Leave One Participant Out

In this strategy, we held out a single participant for validation and trained the model on the other participants. This strategy reflects the cold start case where a new user starts using the health app. This is a generalized approach to model fitting that takes advantage of the existing data set participants.

Leave Week X Out

In Leave Week X Out, we held out a given week for all participants and trained on the rest of the weeks. This strategy evaluates the impact of time-specific segments of data on the prediction. The training phase captures the similarity and variation of the data during different weeks to build the models.

This too is categorized as a general modeling strategy to detect patterns in weekly depressive behavior.

Accumulate Weeks

A sliding window approach was followed in this CV strategy where, for each participant, the model was built with data from weeks *t* to *t*+*n* and tested on week *t*+*n*+1. This strategy examines the feasibility of the personalized ML models using data from individual users and evaluates the impact of longer-term data on prediction accuracy.

Leave One Week One User Instance Out

Here, we trained the models on all the weeks of a participant leaving one of their weeks for testing. This was done for all participants. This method also evaluates the feasibility of the personalized models using each individual user's data on a week-by-week basis without considering the temporal and historical trend.

Baseline Performance

The idea of a baseline was to establish a reference for our accuracy levels. In this study, we operated with a naïve random baseline for our depression score approach and a majority baseline for the depression level change approach. This baseline analysis was carried out for all the CV strategies.

Feature Set–Based Detailed Modeling

Overview

As shown in Figure 1, we used 6 major feature sets: Fitbit, calls, conversations, location, screen, and Wi-Fi. The aggregated data set was used to generate 63 individual data sets that comprised all possible combinations of these 6 feature sets. This approach was used to determine the most effective feature set combination for a specific modeling strategy. Each data set was passed through the modeling strategies and ML algorithms. The process of generating models for the various combinations of data sets is outlined in Textbox 1.

Textbox 1. Process to generate models for the various combinations of data sets.

Process of model generation

- Data set generation: from the aggregated data set that is inclusive of all the feature sets, we generated all possible combinations of the feature sets, including 1-feature sets, 2-feature sets, 3-feature sets, 4-feature sets, 5-feature sets, and 6-feature sets.
- Depression score and change in depression level: we used the Patient Health Questionnaire-9 scores as the ground truth for depression score prediction. To predict the change in depression level, the ground truth was the actual transitions between the depression levels of the participants.
- Machine learning modeling: the derived data sets were all passed through both the universal and personalized modeling strategies.
- Model selection: for depression score prediction and change in depression level both following regression-based approaches, we evaluated and selected the best models based on mean absolute error, mean squared error, root mean squared error, and mean absolute percentage error precision.
- Feature importance: the models that performed best were further analyzed, and their feature importance was calculated. For each combination of sensors and their respective analyses under universal and personalized modeling strategies, the top 10 features were determined. This list of the top 10 features across the combinations was then converted into a frequency chart to help understand the features that had predictive capability.

Feature Importance Calculation

The feature importance in this study was calculated by observing the decrease in node impurity of our tree-based models, which included random forest, AdaBoost, and XGBoost. The impurity for regression tree-based modeling was determined by calculating the variance reduction of the model owing to a feature. Training the tree-based models allowed us to evaluate the contribution of each feature in decreasing the weighted impurity. This decrease in impurity was averaged over the ensemble of trees trained.

Ethics Approval

We obtained approval for this study from the University of Pittsburgh Human Research Protections Office (STUDY18120176). After the participants showed interest in study involvement, they were screened based on the study criteria. The inclusion criteria were that the adolescents be aged 12 to 17.99 years, own an Android or iOS smartphone with access to a data plan, score ≥5 on the PHQ-9 consistent with at least mild symptoms, self-report a previous diagnosis of depression, understand English, and currently reside in the United States. The exclusion criteria were that the adolescents could not have current active suicidal ideation (thoughts with an intent to act on them), a history of a suicide attempt without having received mental health treatment, or a physical deformity or medical reason preventing them from wearing an activity tracker, or be simultaneously participating in a different research study using AWARE. Adolescents meeting the study criteria were offered study participation and, thereafter, if interested, provided their verbal assent, and their parents provided permission. A copy of the consent materials was emailed to all participants for review beforehand.

Results

Overview

In this section, we present the performance of our approaches in predicting depression score and change in depression level in adolescents. The results are the mean values of our runs with the respective approaches. We further show the features that played the most significant role in predicting outcomes for change in depression level. We then report the effect of adding incremental weekly data on the accuracy of depression level prediction. We analyzed the data using both universal and personalized modeling strategies. The study also assessed the impact of missing data on personalized modeling performance. Finally, we conclude this section with a comparative study of classic time-series modeling and a personalized ML model.

Prediction of Depression Score

To understand how sensor features can help in predicting adolescents' depression, we applied regression-based ML algorithms to predict depression scores. The model was compared with a random baseline, and we tested all possible combinations of sensor features and ML algorithms. The evaluation metrics selected were MAE, MSE, MAPE, and RMSE. In particular, we paid close attention to MAE and RMSE. As shown in Table 2, nonlinear algorithms such as the decision trees and AdaBoost performed best. Overall, the personalized models outperformed the universal models in all metrics, in particular MAE and RMSE. The best set of performance metrics recorded was for the Accumulate Weeks personalized strategy. The most optimal model used a 4-feature combination that consisted of Fitbit, calls, screen, and location-based feature sets (MAE=2.39, MSE=10.28, RMSE=2.83, MAPE=0.27). The best results from the personalized models were derived from feature sets that had location, calls, and screen in the combination. This also shows that adding more features does not necessarily yield better results. An RMSE in the range of 2 indicates that our model was able to predict depression scores within 2 scores of test cases.

The results of the regression analysis of depression scores can also be interpreted as levels of depression. This was achieved by segmenting our predictions into intervals of PHQ-9 scores adhering to the established strategy [45]. The confusion matrix of depression levels in Figure 7 was derived based on the depression score predictions to provide more insight. We see that the model is able to best predict levels 2 and 3 most of the time with 89 and 79 correct labels, whereas the other levels, such as 1, 4, and 5, appear to have been predicted 24, 33, and 15 times, respectively. These results support our ground truth distribution, where mild (level 2) and moderate (level 3) depression accounted for most of the samples, followed by moderately severe (level 4), minimal (level 1), and severe (level 5) categorizations.

Table 2. Depression score regression results[a].

	LOPO[b]	LWXO[c]	ACCU[d]	LOWOU[e]
MAE[f] (SD)	4.46 (0.62)	3.43 (0.70)	2.39 (0.10)	2.53 (0.10)
MSE[g] (SD)	30.74 (0.41)	19.0 (0.39)	10.28 (0.21)	11.89 (0.25)
MAPE[h] (SD)	0.55 (0.65)	0.42 (0.52)	0.27 (0.15)	0.29 (0.20)
RMSE[i] (SD)	5.07 (0.71)	4.31 (0.65)	2.83 (0.11)	2.53 (0.17)
Feature set	Fitbit, calls, conversation, screen, location, and Wi-Fi	Calls, conversation, screen, location, and Wi-Fi	Fitbit, calls, screen, and location	Fitbit, calls, conversation, screen, location, and Wi-Fi
ML[j] algorithm	AdaBoost	Random forest	XGBoost	Random forest

[a]The values presented display evaluation metrics for depression score regression models. The best-performing machine learning models were AdaBoost, random forest, and XGBoost.

[b]LOPO: Leave One Participant Out.

[c]LWXO: Leave Week X Out.

[d]ACCU: Accumulate Weeks.

[e]LOWOU: Leave One Week One User Instance Out.

[f]MAE: mean absolute error.

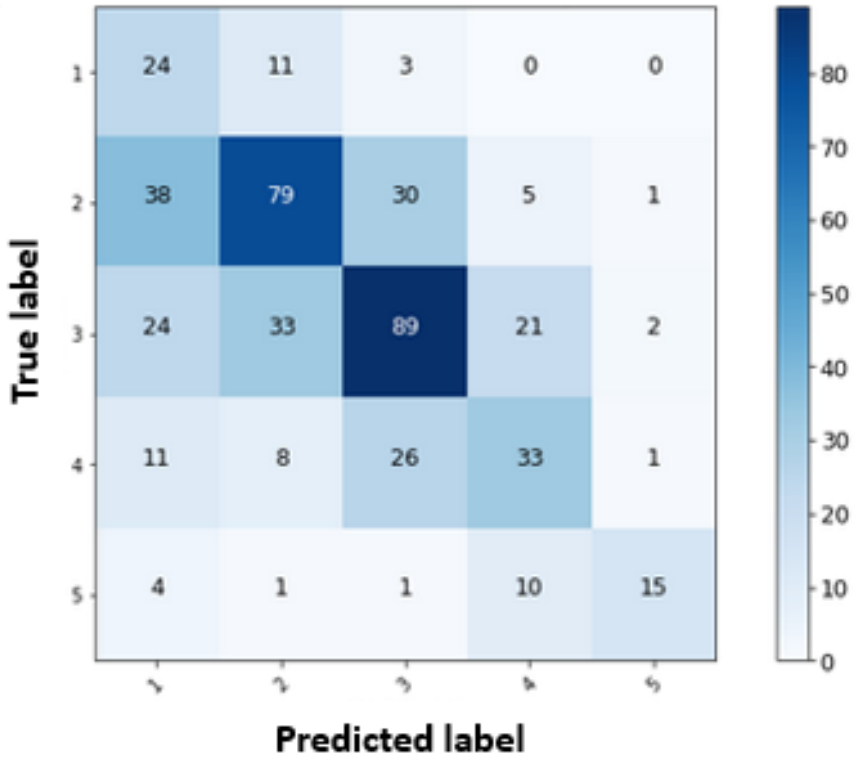
[g]MSE: mean squared error.

[h]MAPE: mean absolute percentage error.

[i]RMSE: root mean squared error.

[j]ML: machine learning.

Figure 7. Confusion matrix of depression levels based on depression score predictions.



Prediction of Change in Depression Level

Table 3 presents the results of predicting change in depression score. In this approach, the change in depression score was calculated between participant weeks as per depression $score_{time}$ – depression $score_{time-1}$. The best-performing models with the lowest MAE were the personalized models Accumulate Weeks (MAE=3.21, MSE=20.13, RMSE=3.86, MAPE=13.69) and Leave One Week One User Instance Out (MAE=3.12, MSE=20.14, RMSE=4.48, MAPE=7.16). Having the ability to predict change within an error margin of –3 to +3 can not only help in determining change in score but also aid in discerning change in levels of depression.

We used the change in depression predictions to create 7 different classes marking changes in levels of depression [45]. The classes (−3, −2, −1, 0, 1, 2, and 3) map the regressed change in depression score to the change in depression level. The signs of the classes represent the rise and fall of depression level, and their values represent the magnitude of change in depression level. Similar to our approach to understanding how depression scores can be interpreted in terms of depression levels, this enabled us to visualize how well our models performed in terms of detecting change in depression score and mapping it to change in depression level. The results from the confusion matrix Figure 8 allowed us to see that the model was able to predict the level jumps (−1, 0, and 1) more accurately than the higher jumps (−3, −2, 2, and 3). This can be explained by the distribution of the observations in the data. Most of the recorded cases witnessed a rise and fall in depression levels by 1 or were at the same level (0) for an extended period. The confusion matrix showed that the true occurrences of large depression level jumps were very rare events.

Table 3. Depression score change regression results[a].

	LOPO[b]	LWXO[c]	ACCU[d]	LOWOU[e]
MAE[f] (SD)	3.28 (0.70)	3.24 (0.67)	3.21 (0.20)	3.12 (0.15)
MSE[g] (SD)	21.35 (0.72)	19.43 (0.63)	20.13 (0.24)	20.14 (0.22)
MAPE[h] (SD)	8.33 (0.55)	15.79 (0.61)	13.69 (0.17)	7.16 (0.20)
RMSE[i] (SD)	4.2 (0.71)	4.26 (0.66)	3.86 (0.18)	4.48 (0.21)
Feature set	Fitbit, calls, conversation, screen, location, and Wi-Fi	Calls, conversation, screen, location, and Wi-Fi	Fitbit, calls, and location	Fitbit, calls, conversation, screen, and location
ML[j] algorithm	AdaBoost	Random forest	XGBoost	Random forest

[a]The values presented display evaluation metrics for depression score regression models. The best-performing machine learning models were AdaBoost, random forest, and XGBoost.

[b]LOPO: Leave One Participant Out.

[c]LWXO: Leave Week X Out.

[d]ACCU: Accumulate Weeks.

[e]LOWOU: Leave One Week One User Instance Out.

[f]MAE: mean absolute error.

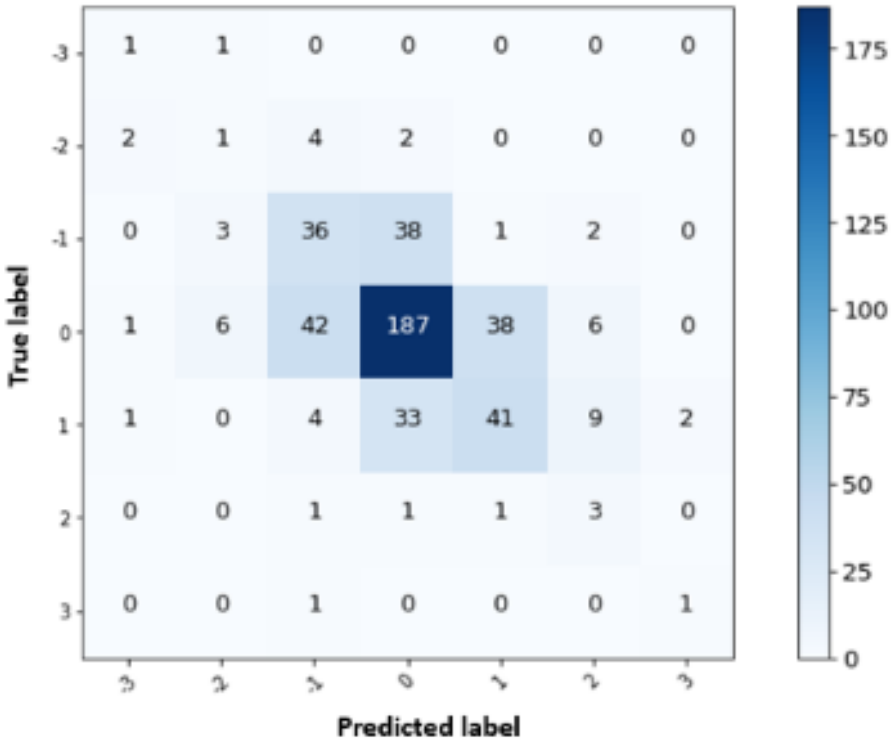
[g]MSE: mean squared error.

[h]MAPE: mean absolute percentage error.

[i]RMSE: root mean squared error.

[j]ML: machine learning.

Figure 8. Confusion matrix for change in depression level into 7 classes (−3, −2, –1, 0, 1, 2, and 3) that represent transitions between higher and lower levels of depression.



Feature Importance Calculation

One of the main advantages of modeling and formulating prediction strategies by extracting features using tree-based approaches is interpretability. In this section, we share the results and provide key insights into the features that were most influential in formulating our ML models. In particular, we chose our depression score prediction results to understand and narrow down the features that played a crucial role in model performance. The results are presented in two parts: (1) plot of the features from the best models based on the frequency of their selection and (2) analysis of the top 10 features with relative importance plotted for each of the personalized and universal modeling strategies.

Most Frequent Features Selected During ML Modeling

Location, calls, and screen were the top 3 feature sets over all modeling strategies. The normalized location entropy and location entropy, which tell us how much time a participant spent at a location, were observed to be most frequently selected during modeling, in particular for both personalized and universal models. The other most frequently selected location features included the outlier time percentage, which is the ratio of time spent in a nonsignificant location divided by the time spent in all locations. Static ratio and number of location transitions were more features that were consistently included in the modeling strategies. Call features were the second most frequent feature in the model. Top call-related features included outgoing calls, in particular Shannon entropy for the duration of all calls, and minimum and mean duration of calls. This was also commensurate with the incoming call features, where, besides the mean, minimum duration of calls also included incoming call count and sum of duration of incoming calls. Frequently selected screen features included first use after unlock, count episode of unlocks, and minimum and maximum duration of screen unlocked. For completeness, we should also mention that conversation, Fitbit, and Wi-Fi followed the aforementioned feature sets.

Figures S1-S4 in Multimedia Appendix 1 show the features that were selected most frequently by Accumulate Weeks, Leave One Week One User Instance Out, Leave Week X Out, and Leave One Participant Out for the best models predicted under them. These plots show the number of times particular features were associated with the best models for all combinations of feature sets under the respective modeling strategy. It is important to note that we modeled up to 6 combinations of sensors and feature sets; therefore, the presence of a feature with a count of 6 indicates that, for all feature combinations tested, that particular feature played a significant role in predictive model building.

Important Features Selected Based on Relative Importance From the Best Depression Score Prediction Models

In the previous section, we presented the results for the most frequently observed features that contributed to the modeling phase. In this section, Figures S5 and S6 in Multimedia Appendix 1 look at the feature importance of the modeling strategies. In particular, we will look at the relative importance among the top 10 features that influenced the respective modeling strategy. Relative importance reflects the importance that the ML algorithm places on a particular feature to form its predictions. Figure S5 in Multimedia Appendix 1 illustrates the feature importance for the Accumulate Weeks (left) and Leave One Week One User Instance Out (right) modeling strategies for depression score prediction. The feature set for Accumulate Weeks that performed best included Fitbit, calls, screen, and location. We see that *screen first unlock* has the maximum importance (0.175), followed by *screen maximum duration unlock* (0.115). They are followed by call features *count of most frequent call types* (0.0754) and *incoming call count* (0.0752).

In the case of Leave One Week One User Instance Out, the 6-sensor combination of Fitbit, calls, conversation, screen, Wi-Fi, and location performed best overall. The best features in this modeling strategy also included *screen first unlock* (0.16) and *screen maximum duration unlock* (0.112). This was followed by *call incoming count of most frequent call types* (0.079) and *screen count episode unlocks* (0.052). An important observation in this result is the similarity of the feature importance of both personalized models. Both modeling strategies selected screen, call, and Fitbit features as important. The results only showed the top 10 features by relative importance; other features such as location followed but had low relative importance.

The universal models are shown in Figure S6 in Multimedia Appendix 1. Both Leave One Participant Out and Leave Week X Out showed the best performance for 5-feature and 6-feature combinations, respectively. We note that both of the generalized approaches ranked the Wi-Fi feature *count of the most scanned access point for a time segment* with high importance (Leave One Participant Out: 0.11; Leave Week X Out: 0.132). Universal models also displayed importance among the screen and Fitbit features. Screen features such as *max duration unlock* (Leave One Participant Out: 0.042; Leave Week X Out: 0.07) and *standard deviation of duration screen unlocked* (Leave One Participant Out: 0.041; Leave Week X Out: 0.038) were also common between the 2 strategies. The Fitbit features of maximum resting heart rate and maximum steps were both selected as important by Leave One Participant Out and Leave Week X Out.

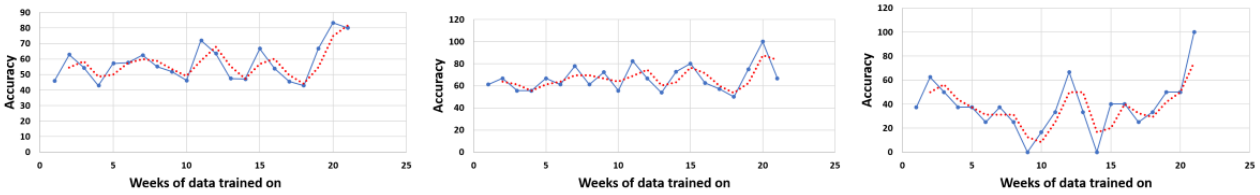
Variation in Accuracy With Increase in Weeks of Data for Accumulated Modeling Strategy

In this section, we present our analysis of how incremental increases in weeks of data affected model accuracy, which was extracted by converting depression score predictions to levels. The results in this section are based on modeling done under the Accumulate Weeks approach. We present 3 plots in Figure 9. The first plot reflects the variation of accuracy based on data from all participants, the second is for participants who showed low SD between their weekly PHQ-9 scores, and the final plot is for participants with high SD between their reported weekly PHQ-9 scores. The blue line represents the accuracy values corresponding to the weeks of data available for modeling, and the red dotted line represents a 2-point moving average.

All 3 plots in Figure 9 show that average accuracy fluctuated between weeks of data available for modeling. Therefore, we looked closely at the trend of the moving average to guide inferences about accuracy variation. In the plot with all participants, looking at the 2-point moving average, we can see that the average accuracy fluctuates between 50% and 60%. The plot with data from participants with a low SD in the PHQ-9 score is higher and can be conservatively stated to be between 60% and 75%. For participants with a high SD in the PHQ-9 score, the moving average shows a significant variation between 20% and 40% accuracy.

We observed that variation in the reported PHQ-9 score by participants contributed to the overall fluctuations. Ideally, it would be expected that, with an increase in data, the accuracy of the models would tend to increase. Although a slight trend upward was observed, the constant rise and fall of participant PHQ-9 scores seemed to affect the accuracy.

Figure 9. Variation in accuracy with increase in weeks of data trained on with a 2-point moving average to map the trend.



(A) Variation in accuracy for all participants

(B) Variation in accuracy for participants with less standard deviation in depression score

(C) Variation in accuracy for participants with large standard deviation in depression score

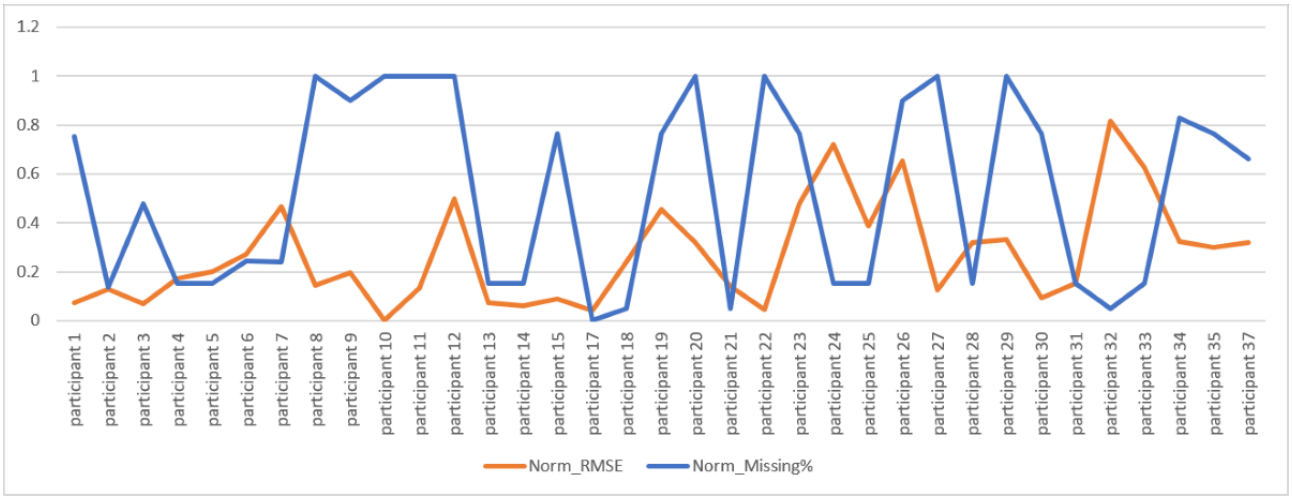
Missingness of Data and the Impact on Accuracy and RMSE

In this section, we explore how missing data affected our accuracy and RMSE values across participants. Figure 10 plots the percentage of missing data for each participant and their accuracy based on personalized modeling (Accumulate Weeks).

We observed 30.89% missing data in the phone-based sensors and nearly 67.74% in Fitbit. Missing Fitbit data were attributed to less than expected adherence to wearing the Fitbit because of reasons relating to regular charging, rash in some participants' cases, and forgetting to wear the device on a regular basis.

The line in blue in Figure 10 is the normalized missing percentage, and the orange line is the normalized RMSE of predicting the depression score. The figure shows how missing data percentage relates to the RMSE value of individual participants for predicting depression score. To analyze both of these values, we normalized them to have the same scale of comparison. Observing a few participants, such as participant 22 (normalized missing percentage: 1; normalized RMSE: 0.04) and participant 24 (normalized missing percentage: 0.15; normalized RMSE: 0.72), we discovered the presence of an inverse relationship between model performance and the amount of missing data.

Figure 10. Missing data percentage (missing%) versus depression score prediction root mean squared error (RMSE)-normalized.



Discussion

Principal Findings

This study presented an in-depth analysis of passively sensed multimodal data collected over a period of 24 weeks from 37 adolescents to predict depression. The collection of data coincided with the COVID-19 outbreak and allowed for the observation of sensor data predictive capability in this scenario. Our models predicted both depression scores and change in the level of depression over weeks. The results showed reasonable improvements compared with the baseline models for both depression score and change in depression level prediction.

We explored universal and personalized modeling strategies. Overall, given the unpredictability of mental health patterns in individuals, personalized models were the most optimal. The Accumulate Weeks modeling approach, which relied on previous windows of sensor observations, achieved an RMSE of 2.83 for depression score predictions and an RMSE of 3.21 for change in depression score prediction. This provides a strong intuition regarding the model's performance. In cases of depression prediction, the model can differ by a score of approximately 2 and, for change in depression score, by a score of approximately 3. This realization of the results points toward the future research and development of more sophisticated personalized predictive modeling to map individual behavioral traits between participants.

Investigating the modeling predictions by segmenting them into depression levels revealed that the model was good at predicting the mild, moderate, and moderately severe levels, whereas minimal and severe levels were difficult to detect because of the less frequent observations in the data collected. In the case of change in depression levels, the models detected decreases and increases with reasonable accuracy when the transitions were −1, 0, and 1. Rare changes such as −3, −2, 2, and 3 were detected with less accuracy. Data imbalance in terms of rare events such as severe changes in depression level, as shown in

this study, can be a subject of further exploration, with possible strategies for synthetic data generation that can imitate sensor readings of participants with sudden or rare changes.

Our study also looked into feature frequency and feature importance. Understanding the features that were selected and highly ranked by optimal-performing models can help in determining what sensors to focus on when analyzing data from passive sensor studies. We see that location, calls, and screen sensor-based features appeared most frequently in the optimal-performing models. Individuals experiencing depressive symptoms tend to move less, which can be captured by location data. Depression also causes participants to reduce their interaction with friends and family, and call-related features can play a role in characterizing this behavior. The screen time of individuals has been seen to be a reflection of mood, as explored in an earlier study [48]. Feature importance helps narrow down the exact features that contributed to modeling. We noticed that, for personalized models, screen time was a strong determining factor that could be a consequence of the COVID-19 lockdown that prompted participants to use their phones with greater frequency. The feature importance presented in this study enabled us to make informed interpretable associations between sensor readings during changes in depression levels or scores. This can propel more research in the direction of more explainable or interpretable model building, especially for mental health–related diseases.

Personalized models performed best in our study of adolescent depression data. Therefore, it was important to understand how personalized modeling, in particular the Accumulate Weeks approach, performed when subjected to incremental data addition as well as looking at how missing values affected model performances. The Accumulate Weeks modeling approach performed better when the variation in the depression scores of the participants was low. By contrast, when the variation in depression scores was high, the accuracy decreased significantly. Exploring the relationship between missing values and the performance metrics of the models allowed us to discover an inverse relationship. This bolsters our understanding that completeness of data can be an important factor in improving model performance. Our experimental analysis of missing data suggests a requirement for strategies to improve the collection of sensor data that can include stronger adherence to protocols by participants or more robust data-generating processes.

Finally, we investigated how autoregressive integrated moving average (ARIMA) models performed in comparison with the ML modeling approaches used. The outcome of the comparison showed that ML models performed relatively better than the classic ARIMA models. However, the ARIMA models were more robust to sudden changes in comparison with the ML models, which were better at predicting smoother transitions. This result leads us to believe that, with possibly larger data sets, a combination of classic time-series approaches and ML-based approaches can be useful for participants with inherent trends or seasonality in their behavior.

Comparison With Previous Research

In the *Related Work* section, we discussed a number of studies on mobile health. In particular, we looked at studies that explored adolescent depression to some degree. Most previous studies were conducted with varied age groups where the adolescent population was either not a part or a very small part of the study [33-38]. However, these studies cannot be considered representative of the adolescent age category. An in-depth review of passive sensing technology for predicting depression [49] mostly focused on college students and adults.

This study focused on adolescents, and all the results are representative of adolescent participants with a previous diagnosis of depression. To the best of our knowledge, this study had the largest sample of adolescents monitored passively to predict depression. In our work, we found location to be one of the most frequent feature sets to be recognized by the ML models. This is in agreement with previous studies of GPS sensor data [33,34] to detect depressive states. They too found a relationship between mobility metrics and depression. The population segment in those studies was restricted to adults. The sensors used were also limited. Other studies that used multimodal sensor data were limited in either participant recruitment or duration [33-37] as well as the population they studied. Our study was an extensive, 24-week–long endeavor with 37 participants being retained for our predictive analysis. A few previous studies on the adolescent population relied on survey-based approaches [40-42]. We differ in relation to them as we strictly based our modeling on multimodal sensor features and did not rely on any direct input from the participants or their parents.

The work by Cao et al [39] was the closest to our study. It was aimed at the adolescent population and used a combination of survey inputs from parents and adolescents besides multimodal features to improve on their accuracy. The differences between the study by Cao et al [39] and our study lie in the type of modeling approaches we used (eg, universal and personalized), the duration of the trial, and the number of participants. In relation to the modeling approaches, Cao et al [39] only performed a universal approach with the best RMSE value leading to 3.70, which combined parents' inputs, steps, GPS, SMS text messages, and calls. We achieved an RMSE of 2.39 based only on the sensor combinations of calls, screen, location, and Fitbit. Their trial lasted 8 weeks and only had 8 participants; this was less compared with our study. Overall, our study explored adolescent depression based on passively sensed data in greater depth owing to both modeling approaches, the study duration, and the participants involved. We also showed how data affected our modeling and compared them with classic techniques such as ARIMA. This information is pertinent in understanding the adolescent population and provides evidence of the type of modeling approaches and features that can generate the best results in predicting depression and change in depression score.

Limitations

Despite the exhaustive modeling approach and strong participant involvement, our study also encountered some limitations. One of the primary limitations of our study was the start of the COVID-19 outbreak that caused potential deviation from the regular behavior of adolescents. Schools were closed and mobility was restricted to the confines of participant houses or

rare outings. Most adolescents were at home and were restricted to television, games, or cell phone use.

We also encountered missing data partially because of the participants' lack of adherence to data syncing and management of the app and partially because of technical issues and difficulty in remote troubleshooting. Missing data are a general concern in passive sensing, which we investigated in our work to show the impact they had on modeling. Data completeness can aid greatly in modeling performance.

Despite our study being one of the studies of longest duration conducted to the best of our knowledge, it can still be categorized as a small data set. A small data set can have an impact on the modeling of rare events; for example, large, sudden jumps in depression scores or extreme depression scores can be hard to track for ML models. Although these measures are anomalies in the data set, perhaps a more focused study on participants exhibiting such traits can be looked into for future directions.

Finally, although our extensive analysis provides useful insights into the feasibility and challenges of using passive sensing for the prediction of adolescents' depression, we emphasize that our study is exploratory and further investigation and more studies are needed to replicate these results.

Conclusions and Future Directions

In this exploratory study, we investigated the feasibility of using passively sensed data for predicting adolescents' depression. We applied universal and personalized ML approaches to predict depression score and change in depression level in adolescents. Our results showed RMSE values of approximately 2 and 3 for the prediction of depression score and for depression change, respectively. This provides confidence in personalized modeling approaches for predicting depression in adolescents. We also investigated the features that models frequently relied on. Features related to screen, call, and location sensors were the most frequent in the optimal models. Our analysis showed better model performance for participants with low variation in depression scores. We also observed that the percentage of missing data of a participant inversely affected the model's performance.

Modeling both change in depression and depression scores can be greatly influential in helping clinicians, parents, and adolescents take preventive measures to intervene in the early worsening of depressive symptoms before entering severe categories. This study will inform the development of an adolescent-facing mobile app with a parent and clinician component to aid in adolescents' self-management and tracking of their mood.

Future research based on our principal findings can help improve mental health prediction. The area of personalized modeling can be used to provide tailored feedback to patients. Rare event prediction in the face of the data imbalance seen in this study should act as an impetus to develop more realistic synthetic data. The feature importance determined can be further explored for other mental illnesses and provide a more interpretable analysis of passive sensing–based studies. Strategies to mitigate missing data for passive sensing will need to balance both participant adherence and modeling strategies that account for missingness. A final promising area of research could be the formulation of an ML model while incorporating a classic time-series approach to account for possible future trends in patients.

Acknowledgments

This study was supported by a grant from the National Institute of Mental Health (1R44MH122067); the National Institute of Mental Health–funded *Center for Enhancing Triage and Utilization for Depression and Emergent Suicidality in Pediatric Primary Care* (P50MH115838); the Center for Behavioral Health, Media, and Technology; and a career development award (1K23MH111922-01A1). Research recruitment was supported by the Clinical and Translational Science Institute at the University of Pittsburgh through the National Institutes of Health Clinical and Translational Science Award program (grant UL1 TR001857).

Conflicts of Interest

SS is a cofounder of the US small business concern (NuRelm, Inc), which was funded through the National Institutes of Health Small Business Innovation Research program to perform this research. Although they have not yet sold a product related to this work, NuRelm hopes to fulfill the Small Business Innovation Research program's goal to disseminate potentially beneficial research through commercialization.

Multimedia Appendix 1
Feature importance.
[DOC File , 221 KB - formative_v6i6e35807_app1.doc]

References

1. Adolescent mental health. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/adolescent-mental-health [accessed 2021-01-02]
2. Beautrais A. Risk factors for suicide and attempted suicide among young people. Aust N Z J Psychiatry 2000 Jun;34(3):420-436 [FREE Full text] [doi: 10.1080/j.1440-1614.2000.00691.x] [Medline: 10881966]

3. 10 leading causes of death, United States. Centers for Disease Control and Prevention. URL: https://wisqars.cdc.gov/data/lcd/home [accessed 2021-01-02]
4. Dick B, Ferguson BJ. Health for the world's adolescents: a second chance in the second decade. J Adolesc Health 2015 Jan;56(1):3-6 [FREE Full text] [doi: 10.1016/j.jadohealth.2014.10.260] [Medline: 25530601]
5. What is depression? American Psychiatric Association. URL: https://www.psychiatry.org/patients-families/depression/what-is-depression [accessed 2021-02-23]
6. Beck AT, Alford BA. Depression: Causes and Treatment. Philadelphia: University of Pennsylvania Press; 2014.
7. Bolton D. Overdiagnosis problems in the DSM-IV and the new DSM-5: can they be resolved by the distress-impairment criterion? Can J Psychiatry 2013 Nov 01;58(11):612-617. [doi: 10.1177/070674371305801106] [Medline: 24246431]
8. Logan D, King CA. Parental identification of depression and mental health service use among depressed adolescents. J Am Acad Child Adolesc Psychiatry 2002 Mar;41(3):296-304 [FREE Full text] [doi: 10.1097/00004583-200203000-00009] [Medline: 11886024]
9. Garnefski N, Legerstee J, Kraaij VV, Van Den Kommer T, Teerds J. Cognitive coping strategies and symptoms of depression and anxiety: a comparison between adolescents and adults. J Adolesc 2002 Dec;25(6):603-611 [FREE Full text] [doi: 10.1006/jado.2002.0507] [Medline: 12490178]
10. Burns JR, Rapee RM. Adolescent mental health literacy: young people's knowledge of depression and help seeking. J Adolesc 2006 Apr;29(2):225-239 [FREE Full text] [doi: 10.1016/j.adolescence.2005.05.004] [Medline: 15996727]
11. Davey CG, McGorry PD. Early intervention for depression in young people: a blind spot in mental health care. Lancet Psychiatry 2019 Mar;6(3):267-272 [FREE Full text] [doi: 10.1016/s2215-0366(18)30292-x]
12. Keenan-Miller D, Hammen CL, Brennan PA. Health outcomes related to early adolescent depression. J Adolesc Health 2007 Sep;41(3):256-262 [FREE Full text] [doi: 10.1016/j.jadohealth.2007.03.015] [Medline: 17707295]
13. Hacker K, Arsenault L, Franco I, Shaligram D, Sidor M, Olfson M, et al. Referral and follow-up after mental health screening in commercially insured adolescents. J Adolesc Health 2014 Jul;55(1):17-23 [FREE Full text] [doi: 10.1016/j.jadohealth.2013.12.012] [Medline: 24525226]
14. Elmquist J, Melton TK, Croarkin P, McClintock SM. A systematic overview of measurement-based care in the treatment of childhood and adolescent depression. J Psychiatr Pract 2010 Jul;16(4):217-234. [doi: 10.1097/01.pra.0000386908.07160.91] [Medline: 20644357]
15. Guo T, Xiang YT, Xiao L, Hu CQ, Chiu HF, Ungvari GS, et al. Measurement-based care versus standard care for major depression: a randomized controlled trial with blind raters. Am J Psychiatry 2015 Oct;172(10):1004-1013 [FREE Full text] [doi: 10.1176/appi.ajp.2015.14050652] [Medline: 26315978]
16. Bánhalmi A, Borbás J, Fidrich M, Bilicki V, Gingl Z, Rudas L. Analysis of a pulse rate variability measurement using a smartphone camera. J Healthc Eng 2018;2018:4038034 [FREE Full text] [doi: 10.1155/2018/4038034] [Medline: 29666670]
17. Nam Y, Kong Y, Reyes B, Reljin N, Chon KH. Monitoring of heart and breathing rates using dual cameras on a smartphone. PLoS One 2016;11(3):e0151013 [FREE Full text] [doi: 10.1371/journal.pone.0151013] [Medline: 26963390]
18. Koenig N, Seeck A, Eckstein J, Mainka A, Huebner T, Voss A, et al. Validation of a new heart rate measurement algorithm for fingertip recording of video signals with smartphones. Telemed J E Health 2016 Aug;22(8):631-636 [FREE Full text] [doi: 10.1089/tmj.2015.0212] [Medline: 26938673]
19. Min JK, Doryab A, Wiese J, Amini S, Zimmerman J, Hong JI. Toss 'n' turn: smartphone as sleep and sleep quality detector. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2014 Presented at: CHI '14: CHI Conference on Human Factors in Computing Systems; Apr 26- May 1, 2014; Toronto Ontario Canada URL: https://doi.org/10.1145/2556288.2557220 [doi: 10.1145/2556288.2557220]
20. Huang K, Ding X, Xu J, Chen G, Ding W. Monitoring sleep and detecting irregular nights through unconstrained smartphone sensing. In: Proceedings of the 2015 IEEE 12th Intl Conf on Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom). 2015 Presented at: 2015 IEEE 12th Intl Conf on Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom); Aug 10-14, 2015; Beijing, China URL: https://doi.org/10.1109/UIC-ATC-ScalCom-CBDCom-IoP.2015.30 [doi: 10.1109/uic-atc-scalcom-cbdcom-iop.2015.30]
21. Montanini L, Sabino N, Spinsante S, Gambi E. Smartphone as unobtrusive sensor for real-time sleep recognition. In: Proceedings of the 2018 IEEE International Conference on Consumer Electronics (ICCE). 2018 Presented at: 2018 IEEE International Conference on Consumer Electronics (ICCE); Jan 12-14, 2018; Las Vegas, NV, USA URL: https://doi.org/10.1109/ICCE.2018.8326220 [doi: 10.1109/icce.2018.8326220]
22. Lin Y, Wong BY, Lin SH, Chiu YC, Pan YC, Lee YH. Development of a mobile application (App) to delineate "digital chronotype" and the effects of delayed chronotype by bedtime smartphone use. J Psychiatr Res 2019 Mar;110:9-15 [FREE Full text] [doi: 10.1016/j.jpsychires.2018.12.012] [Medline: 30611008]
23. Seera M, Loo CK, Lim CP. A hybrid FMM-CART model for human activity recognition. In: Proceedings of the 2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC). 2014 Presented at: 2014 IEEE International Conference

on Systems, Man, and Cybernetics (SMC); Oct 5-8, 2014; San Diego, CA, USA URL: https://doi.org/10.1109/SMC.2014.6973904 [doi: 10.1109/smc.2014.6973904]
24. Doryab A, Togelius J, Bardram J. Activity-aware recommendation for collaborative work in operating rooms. In: Proceedings of the 2012 ACM international conference on Intelligent User Interfaces. 2012 Presented at: IUI '12: 17th International Conference on Intelligent User Interfaces; Feb 14 - 17, 2012; Lisbon Portugal URL: https://doi.org/10.1145/2166966.2167023 [doi: 10.1145/2166966.2167023]
25. Lin M, Lane DN, Mohammod M, Yang X, Lu H, Cardone G, et al. BeWell+: multi-dimensional wellbeing monitoring with community-guided user feedback and energy optimization. In: Proceedings of the conference on Wireless Health. 2012 Presented at: WH '12: Wireless Health 2012; Oct 23 - 25, 2012; San Diego California URL: https://doi.org/10.1145/2448096.2448106 [doi: 10.1145/2448096.2448106]
26. Liu C, Chan CT. Exercise performance measurement with smartphone embedded sensor for well-being management. Int J Environ Res Public Health 2016 Oct 11;13(10):1001 [FREE Full text] [doi: 10.3390/ijerph13101001] [Medline: 27727188]
27. Beiwinkel T, Kindermann S, Maier A, Kerl C, Moock J, Barbian G, et al. Using smartphones to monitor bipolar disorder symptoms: a pilot study. JMIR Ment Health 2016 Jan 06;3(1):e2 [FREE Full text] [doi: 10.2196/mental.4560] [Medline: 26740354]
28. Wang R, Aung MS, Abdullah S, Brian R, Campbell AT, Choudhury T, et al. CrossCheck: toward passive sensing and detection of mental health changes in people with schizophrenia. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2016 Presented at: UbiComp '16: The 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing; Sep 12 - 16, 2016; Heidelberg Germany URL: https://doi.org/10.1145/2971648.2971740 [doi: 10.1145/2971648.2971740]
29. Huckins JF, DaSilva AW, Hedlund EL, Murphy EI, Rogers C, Wang W, et al. Causal factors of anxiety and depression in college students: longitudinal ecological momentary assessment and causal analysis using peter and clark momentary conditional independence. JMIR Ment Health 2020 Jun 10;7(6):e16684 [FREE Full text] [doi: 10.2196/16684] [Medline: 32519971]
30. Costa J, Adams AT, Jung MF, Guimbretière F, Choudhury T. Emotioncheck:: a wearable device to regulate anxiety through false heart rate feedback. GetMobile 2017 Jun;21(2):22-25 [FREE Full text] [doi: 10.1145/3131214.3131222] [Medline: 30923745]
31. Lu H, Frauendorfer D, Rabbi M, Mast MS, Chittaranjan GT, Campbell AT, et al. StressSense: detecting stress in unconstrained acoustic environments using smartphones. In: Proceedings of the 2012 ACM Conference on Ubiquitous Computing. 2012 Presented at: Ubicomp '12: The 2012 ACM Conference on Ubiquitous Computing; Sep 5 - 8, 2012; Pittsburgh Pennsylvania URL: https://doi.org/10.1145/2370216.2370270 [doi: 10.1145/2370216.2370270]
32. Low C, Dey AK, Ferreira D, Kamarck T, Sun W, Bae S, et al. Estimation of symptom severity during chemotherapy from passively sensed data: exploratory study. J Med Internet Res 2017 Dec 19;19(12):e420 [FREE Full text] [doi: 10.2196/jmir.9046] [Medline: 29258977]
33. Canzian L, Musoles M. Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2015 Presented at: UbiComp '15: The 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing; Sep 7 - 11, 2015; Osaka Japan URL: https://doi.org/10.1145/2750858.2805845 [doi: 10.1145/2750858.2805845]
34. Saeb S, Zhang MI, Karr CJ, Schueller SM, Corden ME, Kording KP, et al. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. J Med Internet Res 2015 Jul 15;17(7):e175 [FREE Full text] [doi: 10.2196/jmir.4273] [Medline: 26180009]
35. Doryab A, Min JK, Wiese J, Zimmerman J, Hong J. Detection of behavior change in people with depression. In: Proceedings of the AAAI 14. 2014 Presented at: The Twenty-Eighth AAAI Conference on Artificial Intelligence (AAAI-14); Jul 27–31, 2014; Quebec Canada URL: https://www.aaai.org/ocs/index.php/WS/AAAIW14/paper/viewPaper/8850
36. Xu X, Chikersal P, Doryab A, Villalba DK, Dutcher JM, Tumminia MJ, et al. Leveraging routine behavior and contextually-filtered features for depression detection among college students. Proc ACM Interact Mob Wearable Ubiquitous Technol 2019 Sep 09;3(3):1-33 [FREE Full text] [doi: 10.1145/3351274]
37. Chikersal P, Doryab A, Tumminia M, Villalba DK, Dutcher JM, Liu X, Dey AK. Detecting depression and predicting its onset using longitudinal symptoms captured by passive sensing. ACM Trans Comput Hum Interact 2021 Feb 28;28(1):1-41 [FREE Full text] [doi: 10.1145/3422821]
38. Xu X, Chikersal P, Dutcher JM, Sefidgar YS, Seo W, Tumminia MJ, et al. Leveraging collaborative-filtering for personalized behavior modeling. Proc ACM Interact Mob Wearable Ubiquitous Technol 2021 Mar 19;5(1):1-27 [FREE Full text] [doi: 10.1145/3448107]
39. Cao J, Truong AL, Banu S, Shah AA, Sabharwal A, Moukaddam N. Tracking and predicting depressive symptoms of adolescents using smartphone-based self-reports, parental evaluations, and passive phone sensor data: development and usability study. JMIR Ment Health 2020 Jan 24;7(1):e14045 [FREE Full text] [doi: 10.2196/14045] [Medline: 32012072]
40. Kauer S, Reid SC, Crooke AH, Khor A, Hearps SJ, Jorm AF, et al. Self-monitoring using mobile phones in the early stages of adolescent depression: randomized controlled trial. J Med Internet Res 2012 Jun 25;14(3):e67 [FREE Full text] [doi: 10.2196/jmir.1858] [Medline: 22732135]

41. Jung H, Park HA, Song TM. Ontology-based approach to social data sentiment analysis: detection of adolescent depression signals. J Med Internet Res 2017 Jul 24;19(7):e259 [FREE Full text] [doi: 10.2196/jmir.7452] [Medline: 28739560]
42. Cliffe B, Croker A, Denne M, Smith J, Stallard P. Digital cognitive behavioral therapy for insomnia for adolescents with mental health problems: feasibility open trial. JMIR Ment Health 2020 Mar 03;7(3):e14842 [FREE Full text] [doi: 10.2196/14842] [Medline: 32134720]
43. Maharjan SM, Poudyal A, van Heerden A, Byanjankar P, Thapa A, Islam C, et al. Passive sensing on mobile devices to improve mental health services with adolescent and young mothers in low-resource settings: the role of families in feasibility and acceptability. BMC Med Inform Decis Mak 2021 Apr 07;21(1):117 [FREE Full text] [doi: 10.1186/s12911-021-01473-2] [Medline: 33827552]
44. MacLeod L, Suruliraj B, Gall D, Bessenyei K, Hamm S, Romkey I, et al. A mobile sensing app to monitor youth mental health: observational pilot study. JMIR Mhealth Uhealth 2021 Oct 26;9(10):e20638 [FREE Full text] [doi: 10.2196/20638] [Medline: 34698650]
45. Kroenke K, Spitzer R. The PHQ-9: a new depression diagnostic and severity measure. Psychiatric Annals 2002 Sep;32(9):509-515 [FREE Full text] [doi: 10.3928/0048-5713-20020901-06]
46. Sun Y, Fu Z, Bo Q, Mao Z, Ma X, Wang C. The reliability and validity of PHQ-9 in patients with major depressive disorder in psychiatric hospital. BMC Psychiatry 2020 Sep 29;20(1):474 [FREE Full text] [doi: 10.1186/s12888-020-02885-6] [Medline: 32993604]
47. Pinto-Meza A, Serrano-Blanco A, Peñarrubia MT, Blanco E, Haro JM. Assessing depression in primary care with the PHQ-9: can it be carried out over the telephone? J Gen Intern Med 2005 Aug;20(8):738-742 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.0144.x] [Medline: 16050884]
48. Rosenthal SR, Zhou J, Booth ST. Association between mobile phone screen time and depressive symptoms among college students: a threshold effect. Human Behav Emerg Tech 2021 Feb 18;3(3):432-440 [FREE Full text] [doi: 10.1002/hbe2.256]
49. Meyerhoff J, Liu T, Kording KP, Ungar LH, Kaiser SM, Karr CJ, et al. Evaluation of changes in depression, anxiety, and social anxiety using smartphone sensor features: longitudinal cohort study. J Med Internet Res 2021 Sep 03;23(9):e22844 [FREE Full text] [doi: 10.2196/22844] [Medline: 34477562]

Abbreviations

ARIMA: autoregressive integrated moving average
CV: cross-validation
EDA: exploratory data analysis
MAE: mean absolute error
MAPE: mean absolute percentage error
ML: machine learning
MSE: mean squared error
PHQ-9: Patient Health Questionnaire-9
RMSE: root mean squared error

Edited by A Mavragani; submitted 17.12.21; peer-reviewed by E Sükei, M Kraus; comments to author 03.05.22; revised version received 21.05.22; accepted 22.05.22; published 24.06.22.

Please cite as:
Mullick T, Radovic A, Shaaban S, Doryab A
Predicting Depression in Adolescents Using Mobile and Wearable Sensors: Multimodal Machine Learning–Based Exploratory Study
JMIR Form Res 2022;6(6):e35807
URL: https://formative.jmir.org/2022/6/e35807
doi:10.2196/35807
PMID:35749157

Original Paper

Home Telemonitoring Technology for Patients With Heart Failure: Cost-Consequence Analysis of a Pilot Study

Glory Apantaku[1], MSci; Craig Mitton[1], PhD; Hubert Wong[1], PhD; Kendall Ho[2], MD

[1]School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada

[2]Digital Emergency Medicine, University of British Columbia, Vancouver, BC, Canada

Corresponding Author:
Glory Apantaku, MSci
School of Population and Public Health
University of British Columbia
2206 East Mall
Vancouver, BC, V6T 1Z3
Canada
Phone: 1 604 822 2772
Fax: 1 604 822 4994
Email: glory.apantaku@ubc.ca

Abstract

Background: Heart failure (HF) is a costly health condition and a major public health problem. It is estimated that 2%-3% of the population in developed countries has HF, and the prevalence increases to 8% among patients aged ≥75 years. Home telemonitoring is a form of noninvasive, remote patient monitoring that aims to improve the care and management of patients with chronic HF. Telehealth for Emergency-Community Continuity of Care Connectivity via Home-Telemonitoring (TEC4Home) is a project that implements and evaluates a comprehensive home monitoring protocol designed to support patients with HF as they transition from the emergency department to home.

Objective: The aim of this study is to assess the cost of using the home monitoring platform (TEC4Home) relative to usual care for patients with HF.

Methods: This study is a cost-consequence analysis of the TEC4Home pilot study. The analysis was conducted from a partial societal perspective, including direct and indirect health care costs. The aim is to assess the costs of the home monitoring platform relative to usual care and track costs related to health care utilization during the 90-day postdischarge period.

Results: Economic analysis of the TEC4Home pilot study showed a positive trend in cost savings for patients using TEC4Home. From both the health system perspective (Pre TEC4Home cost per patient: CAD $2924 vs post TEC4Home cost per patient: CAD $1293; *P*=.01) and partial societal perspective (Pre TEC4Home cost per patient: CAD $2411 vs post TEC4Home cost per patient: CAD $1108; *P*=.01), we observed a statistically significant cost saving per patient.

Conclusions: In line with the advantages of conducting an economic analysis alongside a feasibility study, the economic analysis of the TEC4Home pilot study facilitated the piloting of patient questionnaires and informed the methodology for a full clinical trial.

(JMIR Form Res 2022;6(6):e32147) doi:10.2196/32147

KEYWORDS

cost-consequence analysis; feasibility study; pilot study; heart failure; cardiology; cardiovascular disease; economic analysis; telehealth; health care cost; home monitoring; digital monitor; health monitor

Introduction

Heart failure (HF) is a costly health condition and a major public health problem. An estimated 2%-3% of the population in developed countries has HF. The prevalence increases to 8% among patients aged ≥75 years [1]. Although it is the common final stage of many heart diseases, its manifestations can be difficult to diagnose accurately [1]. According to clinical criteria established by the Framingham Heart Study, a diagnosis of HF is confirmed when two major criteria such as elevated jugular venous pressure, pulmonary rales, or a third heart sound are found, or when one major criterion and two minor criteria,

including peripheral edema, dyspnea on exertion, or hepatomegaly, are confirmed [1]. A diagnosis of HF carries substantial risk of morbidity and mortality, despite advances in management.

Home telemonitoring is a form of noninvasive, remote patient monitoring that has gained attention as a promising strategy for improving the care and management of patients with chronic HF. It can be particularly helpful for older adults and those who are frail as well as those at high risk of deterioration [2]. It involves the use of electronic devices and telecommunication technologies (eg, monitoring devices, handheld or wearable technologies, and intelligent sensors) for the digital transmission of physiological and other disease-related data from the patient's home to a health care center providing care and clinical feedback, enabling the collection of clinical data remotely on a regular basis. Using home monitoring technology can result in early detection of clinical decompensation in patients with HF, making it possible to provide timely intervention to prevent mortality events or further deterioration of the patient's condition [2]. Research has shown that for people with HF, structured telephone support and noninvasive home telemonitoring reduces the risk of all-cause mortality and HF-related hospitalizations [3]. These interventions have also been demonstrated to be a major factor in improved self-care behaviors and health-related quality of life and HF knowledge improvements [3].

Telehealth for Emergency-Community Continuity of Care Connectivity via Home-Telemonitoring (TEC4Home) is a project that will implement and evaluate a home monitoring protocol designed to support patients with HF as they transition from the emergency department (ED) to home. The system uses home health monitoring technologies procured by TELUS Health to collect biometric measurements (ie, weight, blood pressure, pulse, oxygen saturation), which feed the monitoring software data to monitor and surveil patient deterioration in an effort to avoid unnecessary ED visits and hospitalizations. In addition, patients are provided a tablet to answer questions on how they feel [4]. TEC4Home has three broad aims. The first aim is to decrease 90-day readmission rates and improve clinical outcomes by increasing the safety and quality of care for patients with HF at home after discharge from the ED [4]. Second, the program aims to help increase patients' engagement and understanding of their condition; the system also aims to improve communication and continuity of care during the transition from ED to home [4]. Finally, TEC4Home aims to achieve a reduction of resource utilization (eg, ED visits and readmissions) to achieve cost savings for the health care system [4].

The aim of this study is to assess the cost of implication of using the home monitoring platform (TEC4Home) relative to usual care and further track costs related to all utilization during the 90-day postdischarge period. Studies have shown that decision makers desire a disaggregated presentation of study costs and outcomes (consequences), which could include changes in survival, quality of life, or indicators of patient satisfaction [5,6]. As such, in this study, a cost-consequences approach was taken. Cost-consequence analysis (CCA) has been defined as an analysis in which costs and effects are calculated but not aggregated into quality-adjusted life years or cost-effectiveness ratios [7]. A CCA that involves comparing the costs and outcomes associated with the home monitoring intervention is more appropriate than a full economic evaluation because this is a pilot study with a limited sample. This form of analysis allows decision makers to compare explicitly the costs associated with usual care and home monitoring technology with the outcomes studied in this pilot. A health system and partial societal perspective was chosen to evaluate the cost-consequence of TEC4Home (ie, direct costs within the health care system and out-of-pocket costs incurred by the patients). This approach was driven by the fact that patient costs and costs outside health care are relevant when it comes to the wider societal impact of this type of technology. This study evaluates the cost implications of the TEC4Home telemonitoring technology on the health system and patients.

Methods

Identification of Outcomes of Interest

The outcomes of interest included in the evaluation are quality of life, mortality, event rates, and costs. Event rates include visits to the ED, general practitioner (GP), or hospital, as well as hospital admissions and length of hospital stay. These outcomes are distinguished by events related to HF and events related to any cause (Textbox 1).

Textbox 1. Disaggregation of the outcome (event rates and costs related to health care utilization).

Event rates

- Number of general practitioner visits
- Number of specialist outpatient visits
- Number of emergency department visits (all-cause)
- Number of hospital admissions and length of hospital stay (all-cause)

Cost components related to health care utilization

- General practitioner visits
- Specialist visits
- Emergency department visits
- Length of hospital stay
- Professional household care, personal care, physiotherapy, and mental health care–related visits (captured as part of out-of-pocket cost)

Cost components related to health care for patients with HF were determined by the TEC4Home trial. Direct costs within health care are derived from those cost components. A distinction was made between costs related to the intervention (including equipment costs, lease costs, and connection fee) and costs related to health care utilization (Textbox 1).

Data Analysis

In this study, utilization includes hospital admissions, ED visits, community family physician visits, and other health provider visits. Utilization was captured through a simple resource utilization questionnaire administered to patients at the time of patient outcome data collection. For the CCA, outcomes are reported in natural units, such as the number of ED visits avoided. Costs are reported in monetary units and consideration is given to costs incurred by the health system and/or the individual patients enrolled in the study. The costs included are those associated with health system resource use, such as ED visits, specialist visits, and nights in hospital.

With each patient serving as his/her own control, we compared health care utilization 90 days before index admission to 90 days posttelemonitoring. The costs of the home monitoring platform relative to usual care were assessed and the study further tracked costs related to all utilization during the 90-day postdischarge period. A 2-tailed paired sample *t* test was used to compare the difference in cost observed between the pre and post periods.

The CCA compared the costs (such as treatment and hospital care) and the consequences (such as health outcomes) of TEC4Home with the standard care patients received before enrolling in the study (which involves clinic visits for clinical examination, assessment of signs and symptoms, assessment of medication use, and provision of self-care instructions).

Results

From October 2016 to June 2017, a total of 519 patients were screened, and 70 patients were enrolled. Patients were excluded if they were unable to complete study procedures, were unable to access a nurse or technology, were having a coronary or structural heart intervention during admission, or had an anticipated survival of less than 90 days. Participants' median age was 75 (range 43-97) years. Complete self-reported health care utilization data from before and after TEC4Home were available for 30 patients; the CCA is based on this data.

The CCA showed a significant reduction in cost associated with length of stay during hospital admission after TEC4Home. With regard to cost associated with ED visits, GP visits, and specialist visits, there was a cost reduction for patients in the home telemonitoring arm; however, this difference was not statistically significant (Table 1).

Additionally, there was a reduction in out-of-pocket costs for patients in the telemonitoring arm; however, this difference was not statistically significant. Patient self-reports on special costs related to their health condition (including drugs, aids to daily living, housekeeping or home care, or transportation to/from medical appointments) showed that patients enrolled in the TEC4Home program saved an average of CAD $118. Note that all dollar values presented in the manuscript are given in Canadian dollars. A currency exchange rate of CAD $1=US $0.78 is applicable.

Table 1. Aggregate health care utilization cost.

	Pre TEC4Home cost (mean), CAD $	Post TEC4Home cost (mean), CAD $	Cost reduction (95% CI) per patient, CAD $	*P* value
Emergency department visit cost[a]	618	262	–87 to 799	.11
General practitioner visit cost[b]	126	129	–52 to 47	.92
Length of stay cost[c]	10,792	3091	3772 to 11,631	<.001
Specialist visit cost[d]	160	132	–72 to 128	.57
Out-of-pocket cost	357	185	–49 to 395	.12

[a]Standard outpatient cost per the Canadian Institute for Health Information: CAD $314.15.

[b]General practitioner visit cost from the Ministry of Health Medical Services Commission payment schedule.

[c]Per diem ward (one night in hospital) per the Canadian Institute for Health Information: CAD $1520.20.

[d]Specialist visit cost from the Ministry of Health Medical Services Commission payment schedule.

From the health system perspective, which was calculated using patient self-report surveys on nights spent in hospital, ED visits, specialist visits, GP visits, other health professional visits, and average cost for TEC4Home (this includes home health monitoring deployment and cost of monitoring nurse), we observed a statistically significant cost saving per patient. In addition, analysis from the partial societal perspective, which included direct and indirect health care costs, showed a statistically significant cost saving per patient. Table 2 shows a breakdown of mean cost from the health system and societal perspective; the relatively wide 95% CIs speak to the small sample size and limited precision in this study.

Table 2. Health care utilization cost from health system and partial societal perspective.

Perspective	Pre TEC4Home cost per patient (CAD $)	Post TEC4Home cost per patient (CAD $)	Cost reduction (95% CI) per patient (CAD $)	*P* value
Health system	2924	1293	1631 (292-2324)	.01
Partial societal	2411	1108	1303 (266-1896)	.01

Discussion

Principal Findings

This CCA showed positive trends in cost savings in the TEC4Home pilot study. Analysis from the health system perspective and the partial societal perspective showed statistically significant cost savings for patients enrolled in the TEC4Home arm. Compared to the 3-month period prior to a patient's index admission, health care utilization in the 3-month period postdischarge was statistically significantly lower for mean length of hospital stay. As a result of their reduced health care utilization, which was mainly driven by reductions in length of hospital stay, patients in the telemonitoring arm cost the health system less than their counterparts in the usual care arm. These patients also had lower out-of-pocket costs than those in the usual care arm.

The development of home telemonitoring technologies can be linked to an improved understanding of the role that early recognition of warning signs of clinical deterioration and responding appropriately in hospital intensive care units play in preventing serious adverse events [8-11]. Home telemonitoring technology targets patients with chronic conditions who have more frequent interactions with the health care system and are thus more exposed to the risk for adverse events. The relatively older skew of the sample in this pilot study reflects that reality. It brings care directly to patients' homes to prevent hospitalization, improve their feelings of safety, and empower them to manage their chronic conditions [11]. The cost of adverse events is a burden on the health care system. The estimated economic burden of preventable adverse events in Canada in 2009-2010 was CAD $397 million. This estimate does not include additional costs incurred by the patients after discharge or costs associated with loss of productivity [12].

Although some studies have measured cost effectiveness, cost utility, and cost benefits of telemonitoring technologies in patients with HF, those studies were done alongside clinical trials. Conducting an economic analysis alongside a feasibility study is often not done, but it is useful in determining the main cost-driving events related to the technology being assessed [13]. It also facilitates the piloting of patient questionnaires to test for clarity and ease of use, ensure pivotal economic data is collected effectively, and estimate completion rates [13]. Furthermore, it provides insight into the sustainability of providing a service like telemonitoring and how such a service can be funded. This study was conducted in recognition of the importance of conducting analysis of relevant data at each point in the development and testing of interventions like home telemonitoring [13].

In line with the advantages of conducting an economic analysis alongside a feasibility study, this study facilitated the piloting of patient questionnaires and informed the methodology for the full clinical trial. The clinical trial has highlighted cost drivers that were not addressed by this pilot study. For example, the clinical trial now includes prescription drug costs using data from PharmaNet, which is a provincewide network that links

all British Columbia pharmacies to a central data system and provides information on every prescription dispensed in community pharmacies. Additionally, the clinical trial includes administrative data collected from all sites to ensure improved accuracy in the measurement of all health utilization variables.

Successful use of telemonitoring technology is not based solely on the efficacy of the technology—rather, it is the result of integration of the technology and existing work practices of patients and clinicians who interact daily with the technology. The presence or absence of successful integration may result in differential technological performance [14-16]. Patient self-care has been identified as a key component of daily HF management [17]. In the adoption of these monitoring technologies, it is important for policy makers to carefully consider how the integration of telemonitoring with existing care management processes may create a need for modifications to existing practices and relations between various health professionals [14]. Policy makers also need to be aware of possible change management costs that come with adopting these technologies. To ensure telemonitoring is cost-effective and clinically effective, it is advisable that there is an effective alignment of proposed technologies with existing practices to facilitate a seamless connection among the various practices, especially in cases where there is a complex organizational setting [14].

Given the rising cost of health care, health planners are looking for alternative methods to provide care to patients that reduce pressure on the health budget while ensuring patients still get high-quality care; one such potential method is telemonitoring. In designing clinical trials to study the effect of these technologies, it is important to ensure the relevant study period is driven by clinical data as this would provide an improved understanding of the role these technologies can play in patient care. These relevant study periods should also drive the cost analyses that are conducted alongside these trials to evaluate the economic implications of adopting these technologies. Additionally, more studies need to adequately evaluate some less obvious costs related to remote monitoring such as database maintenance costs, technical support costs, and possible increases in health care resource use in response to alerts by the monitoring system.

Limitations

There are a number of limitations in this study. The sample size for the TEC4Home feasibility study was relatively small and there may be systematic differences between the patients who were able to sign up for the study and adhere to the monitoring protocol and those who were not (eg, the former may be more willing and able to use technology). However, older patients were included in the pilot study—the average age in the study sample was 74 years. Given that old age is one of the factors associated with less successful self-management, it is important that this intervention aimed at improving self-management was trialed within this population. Another limitation in this study is the absence of controls. Pre-post studies like this pilot study are susceptible to regression to the mean due to the absence of appropriate controls. Regression to the mean highlights the implications of unexplained fluctuations in patient outcomes that are not attributable to the treatment itself; it spotlights the real reasons those fluctuations occur, such as patient adaptation or simple randomness [18]. However, there is a paucity of evidence on the effect of regression to the mean on economic evaluations.

Additionally, this cost analysis is based on self-report of patients and thus is prone to recall bias. The recall period in this pilot study is relatively short to minimize this bias; however, administrative data were available to improve accuracy. This study also does not account for any cost to the patient of using the TEC4Home technology, such as time spent reporting biometric data daily. However, reviews of home monitoring technologies in patients with HF did not provide insights into possible additional costs that patients might incur from using the technology [2]. The follow-up period in this feasibility study is only 90 days and the effect of the intervention on health outcomes and costs will extend long beyond the observation period and these costs will not be captured as part of this analysis. Despite these potential limitations, this study effectively achieved its objectives of detecting potential benefits of the home monitoring technology and providing information on necessary changes and refinements to the larger clinical trial, which can effectively address these limitations, given the proposed 12-month follow-up period and significantly larger sample size. Additionally, the cost analysis contributes to the literature by analyzing direct health care costs incurred by patients that are often ignored.

Conclusion

The CCA showed positive trends in cost savings in the TEC4Home pilot study. Analysis from the health system perspective and the partial societal perspective showed statistically significant cost savings for patients enrolled in the TEC4Home arm. In line with the advantages of conducting an economic analysis alongside a feasibility study, this study facilitated the piloting of patient questionnaires and informed the methodology for the full clinical trial, which is currently underway in British Columbia, Canada.

Acknowledgments

We thank the TEC4Home team for answering our many questions. We also thank the Canadian Institute of Health Research, Michael Smith Foundation for Health Research, BC Ministry of Health, TELUS Health, Vancouver General Hospital and University of British Columbia Hospital Foundation, Chan Sisters Foundation, and a foundation that prefers not to be named for providing financial support for this project. We also acknowledge in-kind support from the Vancouver Coastal Health Authority, Interior Health Authority, Vancouver Division of Family Practice, and patients' physicians.

Conflicts of Interest

None declared.

References

1. Bui AL, Horwich TB, Fonarow GC. Epidemiology and risk profile of heart failure. Nat Rev Cardiol 2011 Jan;8(1):30-41 [FREE Full text] [doi: 10.1038/nrcardio.2010.165] [Medline: 21060326]
2. Kitsiou S, Paré G, Jaana M. Effects of home telemonitoring interventions on patients with chronic heart failure: an overview of systematic reviews. J Med Internet Res 2015 Mar 12;17(3):e63 [FREE Full text] [doi: 10.2196/jmir.4174] [Medline: 25768664]
3. Inglis SC, Clark RA, McAlister FA, Ball J, Lewinter C, Cullington D, et al. Structured telephone support or telemonitoring programmes for patients with chronic heart failure. Cochrane Database Syst Rev 2010 Aug 04(8):CD007228. [doi: 10.1002/14651858.CD007228.pub2] [Medline: 20687083]
4. TEC4Home Healthcare Innovation Community. Supporting Heart Failure Patient Transitions From Acute to Community Care With Home Telemonitoring Technology: A Protocol for a Provincial Randomized Controlled Trial (TEC4Home). JMIR Res Protoc 2016 Dec 18;5(4):e198 [FREE Full text] [doi: 10.2196/resprot.5856] [Medline: 27977002]
5. Coast J. Is economic evaluation in touch with society's health values? BMJ 2004 Nov 20;329(7476):1233-1236 [FREE Full text] [doi: 10.1136/bmj.329.7476.1233] [Medline: 15550430]
6. Drummond M, Brown R, Fendrick AM, Fullerton P, Neumann P, Taylor R, ISPOR Task Force. Use of pharmacoeconomics information--report of the ISPOR Task Force on use of pharmacoeconomic/health economic information in health-care decision making. Value Health 2003;6(4):407-416 [FREE Full text] [doi: 10.1046/j.1524-4733.2003.64245.x] [Medline: 12859580]
7. van Velden ME, Severens JL, Novak A. Economic evaluations of healthcare programmes and decision making: the influence of economic evaluations on different healthcare decision-making levels. Pharmacoeconomics 2005;23(11):1075-1082. [doi: 10.2165/00019053-200523110-00002] [Medline: 16277545]
8. Bleyer AJ, Vidya S, Russell GB, Jones CM, Sujata L, Daeihagh P, et al. Longitudinal analysis of one million vital signs in patients in an academic medical center. Resuscitation 2011 Nov;82(11):1387-1392. [doi: 10.1016/j.resuscitation.2011.06.033] [Medline: 21756971]
9. Celler B, Argha A, Varnfield M, Jayasena R. Patient Adherence to Scheduled Vital Sign Measurements During Home Telemonitoring: Analysis of the Intervention Arm in a Before and After Trial. JMIR Med Inform 2018 Apr 09;6(2):e15 [FREE Full text] [doi: 10.2196/medinform.9200] [Medline: 29631991]
10. Harrison GA, Jacques TC, Kilborn G, McLaws M. The prevalence of recordings of the signs of critical conditions and emergency responses in hospital wards--the SOCCER study. Resuscitation 2005 May;65(2):149-157. [doi: 10.1016/j.resuscitation.2004.11.017] [Medline: 15866394]
11. Kause J, Smith G, Prytherch D, Parr M, Flabouris A, Hillman K, Intensive Care Society (UK), Australian and New Zealand Intensive Care Society Clinical Trials Group. A comparison of antecedents to cardiac arrests, deaths and emergency intensive care admissions in Australia and New Zealand, and the United Kingdom--the ACADEMIA study. Resuscitation 2004 Sep;62(3):275-282. [doi: 10.1016/j.resuscitation.2004.05.016] [Medline: 15325446]
12. Etchells E, Mittmann N, Koo M. The Economics of Patient Safety in Acute Care. Canadian Patient Safety Institute. URL: https://www.patientsafetyinstitute.ca/en/toolsResources/Research/commissionedResearch/EconomicsofPatientSafety/Documents/Economics%20of%20Patient%20Safety%20-%20Acute%20Care%20-%20Final%20Report.pdf [accessed 2022-04-27]
13. Gannon B. The Feasibility Study: A Health Economics Perspective. Global & Regional Health Technology Assessment 2017 Jan 02;4(1):grhta.5000254. [doi: 10.5301/grhta.5000254]
14. Nicolini D. The work to make telemedicine work: a social and articulative view. Soc Sci Med 2006 Jun;62(11):2754-2767. [doi: 10.1016/j.socscimed.2005.11.001] [Medline: 16343724]
15. Christensen JKB. The Emergence and Unfolding of Telemonitoring Practices in Different Healthcare Organizations. Int J Environ Res Public Health 2018 Jan 03;15(1):A [FREE Full text] [doi: 10.3390/ijerph15010061] [Medline: 29301384]
16. Paré G, Moqadem K, Pineau G, St-Hilaire C. Clinical effects of home telemonitoring in the context of diabetes, asthma, heart failure and hypertension: a systematic review. J Med Internet Res 2010 Jun 16;12(2):e21 [FREE Full text] [doi: 10.2196/jmir.1357] [Medline: 20554500]
17. Lycholip E, Thon Aamodt I, Lie I, Šimbelytė T, Puronaitė R, Hillege H, et al. The dynamics of self-care in the course of heart failure management: data from the IN TOUCH study. Patient Prefer Adherence 2018;12:1113-1122 [FREE Full text] [doi: 10.2147/PPA.S162219] [Medline: 29983549]
18. Schilling C, Petrie D, Dowsey M, Choong P, Clarke P. The Impact of Regression to the Mean on Economic Evaluation in Quasi-Experimental Pre-Post Studies: The Example of Total Knee Replacement Using Data from the Osteoarthritis Initiative. Health Econ 2017 Dec;26(12):e35-e51. [doi: 10.1002/hec.3475] [Medline: 28127856]

Abbreviations

CCA: cost-consequence analysis
ED: emergency department
GP: general practitioner
HF: heart failure
TEC4Home: Telehealth for Emergency-Community Continuity of Care Connectivity via Home-Telemonitoring

Edited by A Mavragani; submitted 15.07.21; peer-reviewed by O Joel, A Norris; comments to author 25.08.21; revised version received 14.01.22; accepted 14.03.22; published 02.06.22.

Please cite as:
Apantaku G, Mitton C, Wong H, Ho K
Home Telemonitoring Technology for Patients With Heart Failure: Cost-Consequence Analysis of a Pilot Study
JMIR Form Res 2022;6(6):e32147
URL: https://formative.jmir.org/2022/6/e32147
doi:10.2196/32147
PMID:35653179

Original Paper

Online Knowledge Translation Program Involving Video Games and University Student–Led Tutorials About Cannabis and Psychosis for Black Youth: Mixed Method Feasibility Study

Payal Jani[1], BHSc; Nuri Song[1], BHSc; Erin Artna[2], BSc; Joonsoo Lyeo[1], BHSc; Ashley Assam[1], BHSc; Faith Maelzer[1], BHSc; Andriene Murphy[2]; Angelo Grant[1], BHSc; Suzanne Archie[1], MD

[1]Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada

[2]Department of Psychology, Neurosciences and Behaviour, McMaster University, Hamilton, ON, Canada

Corresponding Author:
Suzanne Archie, MD
Faculty of Health Sciences
McMaster University
East Region Mental Health Services, King Campus - St Joseph's Healthcare Hamilton
2757 King St East
Hamilton, ON, L8N 3K7
Canada
Phone: 1 905 573 4801
Fax: 1 905 573 4802
Email: archies@mcmaster.ca

Abstract

Background: We have piloted a new online knowledge translation (KT) program created to educate youth about cannabis effects, which uniquely focuses on mental health risks for Black youth. Youth are generally unaware of the research linking underage usage and the risk of psychosis. Youth from some Black racialized communities in Ontario may be disproportionately affected and in need of this knowledge.

Objective: Because very little is known about the acceptability and feasibility of programs educating Black youth about cannabis and psychosis risk, we evaluated this KT program, which consists of tutorials facilitated by university students and video games.

Methods: This mixed methods pilot study evaluates the transfer of knowledge about cannabis and psychosis risk before and after the online KT program and, at the same time, explores participant satisfaction with the program and views about underage use. Eligible participants were youth 16-19 years of age of Black African or Caribbean descent. Trained undergraduate students from McMaster University administered a quiz (psychosis and cannabis test; PCT) to evaluate knowledge before and after the KT program. After playing the psychoeducational video games, participants attended two tutorial group sessions led by undergraduate students. The undergraduate students facilitated the online tutorials about cannabis and psychosis. The tutorials augmented the educational content embedded within the gameplay: participants discussed what they learned from the video games and their understanding of psychosis and the effects of cannabis. In addition, undergraduate students qualitatively analyzed the tutorial discussions for themes, and the prequiz and postquiz scores were analyzed for significant differences in scores.

Results: A total of 9 Black youth were recruited and completed this pilot study. The mean PCT scores were 5.67 (SD 1.7) and 7.78 (SD 1.8) before and after the KT program, respectively. There was a significant improvement in scores (P<.05) post-KT program. Thematic analysis of the facilitated tutorials revealed three major themes: video game satisfaction, marijuana and psychosis literacy, and help-seeking awareness. Overall, participants showed an increased awareness and understanding of the subject matter after the gameplay and tutorial intervention.

Conclusions: When supplemented with tutorial sessions, the Back to Reality Series shows promise for addressing the gap in knowledge about cannabis and psychosis, and the results provide preliminary evidence that the games appeal to Black youth.

(JMIR Form Res 2022;6(6):e33693) doi:10.2196/33693

KEYWORDS

knowledge translation; Black youth; video game; psychosis; cannabis use; knowledge; young adult; race; demographic; minority; gaming; mental health; drug; cannabis; acceptability; feasibility; risk

Introduction

Background

Before its legalization for nonmedical use in 2018, cannabis was the most widely used illicit substance in Canada. Among advanced economies, Canada has a relatively high rate of adolescent cannabis use [1], as high as 28% among youths between 15 and 19 years old [2]. Given policies to restrict cannabis use among underage youth, one might expect greater public education geared toward youth. In a 2017 Canada-wide survey, less than 50% of youth could identify the mental health effects of cannabis [3,4]. Gaps in knowledge about risks of underage use have also been noted in the United States [4].

Even though the majority of healthy adult cannabis users do not experience health risks from use [5], studies have found a significant relationship between underage cannabis consumption and the onset of subsequent psychosis [6-8]. Several risk factors influence this relationship: the age of onset of regular use [8], frequency of use [8], genetic vulnerability for schizophrenia [9,10], previous psychosis symptoms [11], genetic risk for schizophrenia, and the delta-9-tetrahydrocannabinol (THC) content or potency of the cannabis consumed [11]. THC is the psychoactive component of cannabis associated with addiction and hallucinatory experiences [12]. The higher the THC content, the greater the risk of experiencing these effects [8]. Underage cannabis use may pose additional barriers for youth from Black racialized communities.

There are few public health tools about cannabis use and its mental health effects (eg, Canada's Lower-Risk Cannabis Use Guidelines [13]), but generally, public education about the mental health effects of cannabis for youth is lacking, particularly in Black communities. Black youth have experienced disproportionate rates of criminalization, stigma, and negative stereotypes associated with cannabis use [14]. Credible knowledge delivered in a culturally safe manner is needed for this population. A recent Ontario study suggests that people from Black Caribbean communities versus Canadian-born respondents had higher rates of past-year cannabis use (odds ratio [OR] 1.70, 95% CI 1.04-2.79; $P<.01$). Furthermore, this group's risk was also higher (OR 2.76, CI 1.24-6.12; $P<.05$) for problematic cannabis use—defined as a pattern of use associated with harm, abuse, or dependence [15]. In contrast, for other Black ethnic groups, the OR was lower for people from the Black African group for past-year use and was not significantly different from Canadian-born respondents (OR 0.68, 95% CI 0.35-1.31) [15]. Black racialized communities do not represent a monolith. Interventions designed explicitly for Black youth are rare [16], and even these could benefit from innovative strategies that appeal to diverse ethnic groups within Black communities.

Interactive Tutorials and Peer Modelling

Interactive tutorials are a promising educational method compared to traditional lecture-based or textbook instructions [17]. Active learning techniques are associated with higher skill acquisition and improved motivation for learning (eg, use of interactive video games, informal games, and in-class time devoted to discussions about a case) [17]. University student–led tutorials for more junior learners offer peer role modeling and may also enhance professional and communication skills [18]. The undergraduates learn to explain the scientific knowledge, and the participants are encouraged to engage in critical thinking about the science [18]. Combining multimedia strategies such as animation with community-based programming offered by schools or family physician's offices has augmented behavioral change, even reducing illicit drug use [19].

Serious Video Games

Online digital video game technology offers engrossing platforms for youth to receive and integrate mental health information [20]. Immersive and interactive story lines have been shown to promote health-related behavioral change and facilitate deep learning [21,22]. Therapeutic video games are an emerging area in the health care field because of their capacity to simulate real-life symptoms and treatment [23]. Video game technology has improved outcomes for US Army veterans with posttraumatic stress disorder [24]. Sparx, a video game that uses avatars to deliver cognitive behavioral therapy (CBT), is effective for depression using a randomized control design [25,26]. Reductions in paranoid ideations and anxiety have been achieved for patients with psychosis compared to treatment as usual, using immersive virtual reality CBT for patients [27]. Nonetheless, media campaigns that warn youth about the harms of substance use, particularly tobacco or illicit drugs, have yielded mixed results [28].

The Back to Reality Video Game Series

The Back to Reality Video Game Series (the SERIES), a knowledge translation (KT) product, was created to translate messages inspired by research on cannabis use [29,30] and pathways to care for a first episode of psychosis for young people of Black African and Caribbean descent [31,32]. The games were produced with input from an integrated KT community of Black youth, students, young people with lived experiences, family members, game designers, and researchers. It was created using principles of serious game design methodology for effective game design in education [33]. Interactivity helps players explore the potential benefits, harms, and emotional, social, or psychiatric consequences of regular cannabis use. It depicts Harry and his group of friends from diverse ethnic backgrounds exploring the potential positive and negative consequences of cannabis use. Harry is an 18-year-old second-generation Canadian youth of Jamaican descent who develops psychosis after regular cannabis use and enters a virtual mental health system as a result. It "shows" rather than "tells," using a medium appropriate for the audience. Family physicians

thought the SERIES met their needs to educate young people about cannabis use, mental health, and addiction services [34]. They expressed a willingness to offer it to young people with mental health and addiction issues and their families [34].

These studies support the use of serious video games and interactive learning strategies to address mental health and addictions issues. Nevertheless, the existing literature lacks data on interventions designed to help youth from Black communities understand the mental health impacts of cannabis. Black youth from Caribbean communities may be more at risk for problematic cannabis use [15] and more vulnerable to the impact of stigmatization and criminalization [14]. However, very little is known about how to translate research knowledge for this population. In this pilot study, Black youth were exposed to video games and university student–led tutorials to evaluate the feasibility and acceptability of this KT program about cannabis and psychosis. The participants are expected to acquire relevant knowledge about underage cannabis use and the risk of psychosis. This paper explores themes emerging from the tutorial sessions about participant satisfaction with the KT program. This evidence could spur future randomized control trials on the effectiveness of educational and online digital strategies for increasing learning outcomes about the mental health impacts of cannabis.

Methods

Overview

This project collected data on a new KT program involving psychoeducational video games and tutorials designed to educate Black African and Caribbean youth about the relationship between cannabis and psychosis. Our objective was to explore the extent to which this online KT program could transfer health information to youth—a challenging population to engage in medical literacy—and assess user satisfaction. This pilot project was based on a partnership involving McMaster University and the Free for All Foundation, a community charitable organization in Ontario that serves Black youth and their families. It is a grassroots community-based program that provides support so that Black youth can fulfill their potential.

The study used qualitative and quantitative data collected before, during, and after two tutorial sessions facilitated by undergraduate students. The primary variable was changes in scores on a knowledge test, the psychosis and cannabis test (PCT) quiz. The training was conducted from September to December 2020, and the recruitment, data collection, and analysis occurred between January and April 2021.

A group of 10 undergraduate students at McMaster University participated in a research-based thesis course under the supervision of the senior author. Over 4 months, these undergraduate students studied the research relevant to the association between youth cannabis use and psychosis, using the Back to Reality Series as the inspiration for their learning objectives. They were taught how to facilitate tutorial sessions. The undergraduate students were trained in the proper conduct of data collection, transcribing the data, analyzing, coding procedures, generating themes, and writing up the report. This manuscript is based on input from many of the undergraduate students supervised by the senior author.

Convenience sampling was used to recruit Black youth from the community organization. The inclusion criteria were male, female, or transgender Black youth registered at the community organization and aged between 16 and 19 years. Youth were excluded from the study if they had a known seizure disorder or a video game addiction. Before their inclusion, each participant was administered a screening tool to assess their eligibility.

Ethical Considerations

This study was approved by the Hamilton Integrated Review Ethics Board (11181), and each participant provided written informed consent for participation.

Interventions

The online KT program consisted of two parts: tutorial sessions facilitated by undergraduate students and the Back to Reality Series video games.

Tutorial Sessions

There were 3 tutorial groups. Each group underwent two online tutorial sessions, each led by 3-4 undergraduate students. Tutorials lasted 60 minutes each and were conducted over Zoom. The undergraduate students translated the scientific knowledge about cannabis and psychosis they had acquired the semester before into simplified teaching content for the participants. The learning objectives were to increase the participants' awareness of the benefits and harms of cannabis. The goal was to help participants understand why underage cannabis use carried a greater risk of adverse consequences than adult-onset use, based on scientific research. Other learning objectives were as follows:

- Understand psychosis disorders and cannabis use disorder
- Appreciate factors mediating the risks associated with underage cannabis use and psychosis, including the following:
    - Age of first use
    - Potency of marijuana
    - Genetic vulnerability
    - Choice of products
    - Pathways to care for marijuana use disorder and psychosis

The learning objectives were inspired by the narratives explored within the video game play, and the undergraduate students conferred with each other to ensure similar content was covered across tutorial groups. All participants played the SERIES, establishing a common ground for educational content. Active learning strategies such as interactive games, discussions, and problem-based learning incorporating narratives from the gameplay were incorporated into the tutorial sessions facilitated by the undergraduate students. Participants were able to discuss their experience playing video games and have their questions answered.

Back to Reality Series

The SERIES consists of 3 video games: Harry's Journey, Harry's Journal, and Harry's PathwaysToCare map. Harry's

Journey imparts the story of Harry—as he considers whether or not to seek help for his psychosis and cannabis use (Figure 1). Harry's Journal delivers experiential knowledge about major psychiatric symptoms. The PathwaysToCare Map displays 3D replicas of youth mental health and addictions services so that young people can learn about the health care system and how to access care. McMaster University owns the intellectual property.

Figure 1. Harry's Journey main menu scene.

In a demonstration project [35], 20 undergraduate university students played the SERIES and a control game. All participants played both games, but the order was randomized within the software to determine which game was played first. The participants completed a quiz after playing each game. Playing the control game first, followed by the SERIES, led to significant increases in the quiz scores (P=.005). Participants randomized to playing the SERIES first, followed by the control game, had no significant changes in quiz scores, suggesting the SERIES was effective in transferring knowledge, but the control game was not.

The SERIES has been pilot tested among youth aged between 16 and 19 years and experiencing homelessness: 45% (25/55) were neither in school or working, 24% (13/55) reported psychosis experiences, and 13% (7/55) were of Black African/Caribbean descent. Furthermore, 88% (48/55) of the sample had a lifetime use of cannabis, with an average cannabis use onset of 13 years. This pilot study involving 55 youth demonstrated a significant mean knowledge test score advantage for participants playing the SERIES (54%; P=.02; mean 6.7, SD 1.7) compared to those playing the control game (mean 5.5, SD 2.0) [36]. The SERIES yielded a 22% improvement in test scores compared to the control game. The majority of participants (49/55, 90%) playing the SERIES enjoyed it versus 75% (41/55) of those playing the control game. Furthermore, the qualitative analysis revealed that Harry's mental health experiences resonated with them, with participants saying "I felt like I knew what he was going through" and "Um, a lot of people make it run their life, and they skip school…like weed of all things their top priority."

During preliminary testing, 10 participants aged between 17 and 30 years with a first episode of psychosis felt that Harry's Journey realistically portrayed psychosis experiences, and they enjoyed playing it [37].

Measures

Demographic data were collected on age, gender, and highest academic achievement.

PCT Quiz

The PCT quiz is a test of knowledge for participants. It does not involve clinical judgment or measurement of a clinical condition. The PCT quiz was constructed and validated to examine whether the Back to Reality Series delivers knowledge about the relationship between psychosis and cannabis (Multimedia Appendix 1). It consists of 10 multiple-choice knowledge questions worth 1 point each. The questions were reviewed, pilot tested, and revised by a 15-year-old high school student and a psychiatry resident who were part of a KT community that constructed and evaluated the video games. Items were generated from concepts arising out of the literature review concerning the relationship. When administering the quiz, student researchers read the questions aloud to reduce the impact of reading literacy levels. It takes 10 minutes to complete the quiz.

The PCT quiz has produced consistent results supporting its reliability—replicating significant differences in knowledge acquisition scores after exposure to the SERIES across 3 small-scale demonstration projects. In 2020, a group of 20 McMaster University undergraduate students aged 17-22 years played two games, Morpheus Spell (a control game) and the SERIES [35]. The order in which the students played the game was randomized, but they all played both video games and completed the same two quizzes after playing each game. The participants were administered quiz 1 (a more challenging

version of the PCT quiz for older youth) and quiz 2 (the PCT quiz). Participants had significant increases in scores on both quiz 1 and the PCT quiz when they played Morpheus Spell first, followed by the SERIES. Both quiz 1 and the PCT quiz produced significant increases in scores when participants played Morpheus Spell first, followed by the SERIES: on quiz 1 the scores were 5.44 (SD 1.51) versus 6.78 (SD 1.48, $P<.03$), and for the PCT quiz the scores were 6.67 (SD1.41) versus 8.22 (SD 1.30, $P<.005$). Statistical significance was not achieved for either quiz when participants played the SERIES first, followed by the control game (quiz 1: $P=.17$ and PCT: $P=.28$). The control game did not transfer any knowledge relevant to cannabis use or psychosis. This latter finding validates the use of the PCT quiz to test knowledge acquisition.

The PCT quiz was pilot tested on 10 clients (aged 17-30 years) with a first episode of psychosis using a pre-post design [37]. The pre-PCT quiz mean score was 6.5 (SD 1.3) versus the post-PCT quiz mean score of 7.7 (SD 1.4), revealing a significant increase (18% improvement) in posttest scores ($P<.01$), supporting its face validity [37]. The PCT quiz was also used with 55 youth experiencing homelessness who were randomized to either the SERIES or the control game first, as described above [38], revealing a statistically significant improvement (>18% difference) in scores ($P<.05$).

Postgameplay Survey

Participants were given a postplay survey consisting of 10 questions that gathered information on how often they play video games, which devices they use, their level of comfort while playing the Back to Reality Series, whether they enjoyed different aspects of the game (recommend game to a friend; thought game was youth-friendly; and enjoyed the story, the music, the graphics, the basketball minigame, and the game as a whole). The satisfaction questions were scored yes (1) or no (0). The survey was adapted from a study examining the acceptability and satisfaction with a video game promoting messages and actions concerning antenatal care [39].

Procedures

Participants took part in 3 online zoom sessions conducted by undergraduate students. During visit 1, the participants answered structured demographic questions and open-ended questions about their understanding of cannabis and psychosis. Undergraduate students administered the PCT quiz to establish the participants' knowledge base. Next, undergraduate students shared their screens with participants who played Harry's Journey over Zoom. The subsequent two sessions involved tutorials led by the university students. All tutorial discussions were recorded over Zoom. After the first tutorial (visit 2), participants played the remaining games—Harry's Journey and the PathwaysToCare map. After the second tutorial (visit 3), participants completed the postplay survey to assess their satisfaction with gameplay. Participants answered qualitative questions about their gameplay experience and their views on cannabis. The PCT quiz was readministered to assess changes in their level of knowledge.

All responses were recorded and transcribed by the undergraduate students. Identifiers were removed from each transcript and amalgamated into one encrypted document for thematic analysis by undergraduate students and the senior author.

Data Analysis

Statistical Analysis

Chi-square and *t* tests were used to analyze the demographic and pre-post gameplay survey. A mean and standard deviation were calculated for participants' PCT scores before and after the gameplay and tutorial intervention. A 2-tailed paired *t* test was then conducted using the before and after PCT scores to determine whether there was a significant change in scores. A *P* value <.05 was deemed significant for this statistical test.

Qualitative Analysis

The tutorial sessions for all groups were transcribed and consolidated into one document for all the undergraduate students to analyze. Each undergraduate student independently analyzed the data corpus, using the thematic analytic method espoused by Braun and Clarke [40] to identify patterns and meanings from data extracts and initial codes. Each student independently coded the data into themes and a thematic map. Two authors (PJ and SA) reviewed these codes and data extracts and selected the most poignant. The final descriptive thematic map organizes and highlights the prominent themes and subthemes.

Results

Overview

There were a total of 9 participants. Table 1 outlines the demographic characteristics of the 9 Black participants. The majority of participants (7/9, 77%) were male and attending or recently completed high school. There was complete attendance for all 3 visits.

Table 1. Participant demographic data (N=9).

Variable	Values
Race, Black, n (%)	9 (100)
Age (years), mean (SD)	17.56 (1.17)
Gender, n (%)	
Male	7 (78)
Female	2 (22)
Educational enrollment, n (%)	
High school	5 (56)
Postsecondary	2 (22)
Not in school[a]	2 (22)

[a]Completed high school.

PCT Quiz

Table 2 shows a significant difference in PCT quiz scores before (mean 5.67, SD 1.66) versus the post-PCT quiz score (mean 7.78, SD 1.79) after the KT program (*P*=.01). Overall, 7 participants (77%) showed improvement in their PCT scores, and 2 (22%) participants had scores that remained unchanged.

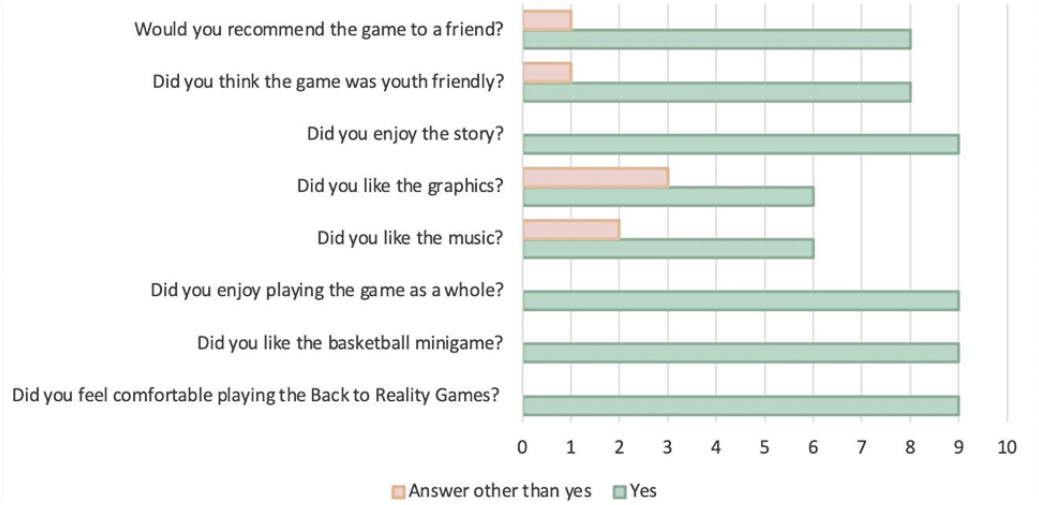
Table 2. Knowledge translation program scores before and after participation.

Statistic	Values
Score, mean (SD)	
Before	5.67 (1.66)
After	7.78 (1.79)
95% CI	
Before	4.58-6.75
After	6.61-8.95
P value	.01

Postgameplay Survey

The majority of participants (5/9, 56%) played video games at a frequency of less than once per month. Smartphones were the most popular device used for gaming. Figure 2 outlines satisfaction with the Back to Reality Series based on the postgameplay survey. Participants unanimously reported they enjoyed the story, gameplay experience, and basketball minigame. The least satisfactory elements were the graphics and the music. The criticisms were reported by participants who played video games "almost every day" or "at least once a week."

Figure 2. Postplay survey of gameplay satisfaction.

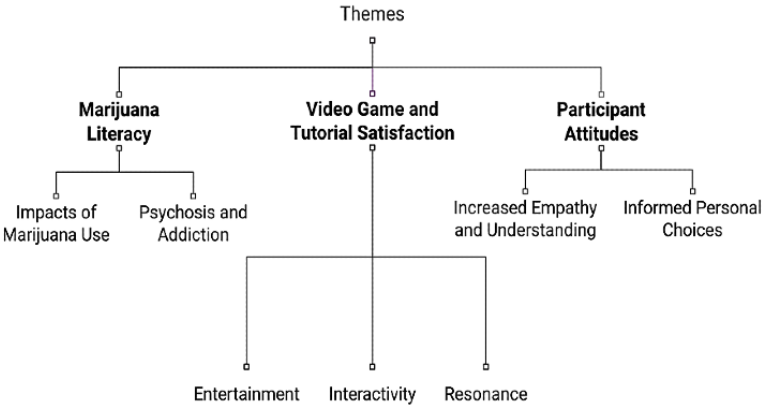


Qualitative Analysis

Tutorials supported further deliberations about video games in an environment where participants felt supported by postsecondary students. Participants were given a chance to hear from other youth and formulate answers to questions. Participants were often seen learning from one another, adding to each other's answers, and engaging in friendly competition during interactive activities.

Discussions from the transcribed tutorials and qualitative interviews were classified into three major themes: (1) video game and tutorial satisfaction, (2) cannabis and psychosis literacy, and (3) participant attitudes. Under each central theme, 2-3 subthemes were identified, shown in Figure 3.

Figure 3. Thematic analysis map of tutorial discussions and video game feedback.



Video Game and Tutorial Satisfaction

The first theme encompassed participants' opinions about playing the SERIES and attending the educational tutorials. Several elements of the intervention were discussed; however, participants' comments primarily reflected the degree of entertainment and interactivity they experienced when playing the video game. Furthermore, participants identified with the main character and experienced the narrative as authentic.

Entertainment

When asked to characterize their experiences with the game, many participants commented on the dualities of entertainment and education, suggesting entertainment increased engagement in their learning. Additionally, participants frequently used terms such as "fun," "enjoying," "entertaining," and "like" to describe their experience with the intervention, showcasing a positive attitude toward it (eg, "It was actually fun. I enjoyed the game, and I also learned a lot too… It was a great experience").

In the postgameplay survey, 100% (9/9) of participants said that they enjoyed playing the Back to Reality Series as a whole. The surprise element of the mental health experiences helped, particularly when the plot contrasted with their expectations and knowledge.

> *Yeah, I would because, uh it's a visual way to represent what someone in that situation would be going through, and I feel like maybe someone who does use marijuana, then it can be a sort of visual way for them to see someone who its affecting negatively. Yeah, this would be awesome.*

Interactivity

The KT program integrated interactivity heavily through gameplay involving the video games and open-ended discussions during tutorials. This interactivity was consistently noted by participants, who suggested its importance in their understanding of concepts (eg, "I liked the interacting [tutorial] games that we did, it was fun. Because we were playing games but also learning").

Additionally, the first-person point of view of the video game may have led to increased personal investment in the outcomes of decisions. Since these outcomes revealed information about cannabis-associated risks, this engagement may have facilitated learning.

> *In terms of the game, I really like the fact I was able to control the decisions that Harry made cause then I can see how based on those decisions how it affected his life. And yeah...my biggest takeaway was really the decision-making aspect.*

Resonance

The story line in the SERIES included many elements of teenage life, such as school, relationships, and families. As a result, many participants reported feeling more connected to the intervention, increasing engagement. Moreover, some participants discussed having friends who used cannabis and reported that the content was meaningful to their lives (eg, "I guess the fact that like Harry lived a realistic teenage lifestyle that like a lot of youth can relate to and that made it a lot more meaningful cause it felt more relevant").

When asked if they would recommend the game to a friend, one participant stated, "Yeah…Because they actually learn and see kind of like, um, like a reflection of themselves."

Cannabis and Psychosis Literacy

The transcribed data set revealed a significant improvement in cannabis and psychosis-related knowledge among participants postintervention. All 9 participants described their participation in the intervention as educational or informative. Participants

primarily acquired knowledge about the impacts of cannabis use, psychosis, and addiction.

Impacts of Cannabis Use

Thematic analysis revealed that participants had a sound understanding of addiction before starting the KT program (one participant said "Addiction, like it's not a hobby, it's more extensive than a hobby like it's something you do commonly because you can't stop").

However, mental health risks were a prevalent but novel point of discussion in their postintervention interviews, suggesting newly emerged awareness achieved through their participation. Participants comment on the gravity of these mental health risks, adding nuance to this awareness: "Looking at it through like a mental lens, it could lead to like, dependency and addiction, um, which is also like a very big, um, mental health problem."

Similarly, preintervention transcripts showed participants primarily discussing the social benefits of cannabis use and neglecting its social harms. Postintervention, there was a substantial consideration for these harms, with participants noting how in the Back to Reality Series, Harry's relationships with his girlfriend, mother, and friends were negatively impacted due to his cannabis use and emerging psychosis (one participant noted "It might ruin friendships…[Harry] was paranoid with his social life he thought everybody was out to get him"). This direct reference to the video game resulting in their increased knowledge showcases its effectiveness as a multimethod educational tool.

Enhanced Psychosis and Addiction Knowledge

Prior to gameplay and tutorial intervention, 66.7% (6/9) of participants indicated that they did not know or were unsure of the meaning of psychosis.

> *I kind of learned a lot from it and the effects [marijuana] has. And also, like what psychosis is, because before when the game started, I didn't know anything about psychosis.*

Post-KT, participants were able to define psychosis and even distinguish between delusions and hallucinations, showing a nuanced understanding of the subject. There was a new awareness of the mental health risks of cannabis. Participants showed a strong understanding of the risk factors leading to cannabis addiction, psychosis, and the signs of addiction among cannabis users.

> *Someone that has psychosis will have lots of hallucinations—auditory, visual—and it will affect your perception of like reality.*
>
> *I think it makes sense that like early-onset…would increase your risk of psychosis and addiction.*
>
> *A sign that a youth is using marijuana to excess is that] they're losing touch with their friends, isolating themselves, staying at home, not talking to anybody…being antisocial.*

Participant Attitudes

As a result of their newfound knowledge, a shift in attitude was seen among participants. Their postintervention comments reflected an increased sense of empathy and understanding for those suffering from cannabis addiction and seeking help. Additionally, participants became more self-reflective regarding their own cannabis use choices.

Increased Empathy and Understanding

Many participants expressed how their understanding of cannabis use became more balanced after playing the Back to Reality Series and attending the tutorial sessions. Some acknowledged how this challenged their previous misconceptions regarding cannabis users and allowed them to minimize their judgments toward this population.

> *Before coming into here I already had a bad connotation of marijuana because I just viewed it as bad…Then I played the game and came to these Zoom meetings with you guys, and I learned even more about why it's bad and what it can do for you, both the good and the bad.*

Additionally, despite concerns about stigma, the importance of help-seeking for symptoms was a key message identified by some participants. Participants linked getting professional help to positive mental health consequences as opposed to negative consequences associated with delays. Many credited the video games with normalizing the process of seeking help, challenging pre-existing hesitancy they had about pathways to care for mental health and addiction issues.

> *Getting help is not something to be embarrassed about, it's something that can actually help you…And the experience that [Harry] didn't...he got worse when he didn't get help.*
>
> *I guess when he got help at the support center because they didn't come off as it was his fault, they came off as...we address the situation, how can we help you now.*

Informed Personal Choices

Four participants (44%) explicitly weighed out the pros and cons learned from the intervention to inform their decision-making (for their own and a hypothetical friend's cannabis use). By linking knowledge to personal conclusions, the intervention is seen to impact their choices.

> *I personally think it's…not necessary to be using [marijuana], especially considering the risks. I just think the cons outweigh the pros.*
>
> *All the side effects...all those people who were originally affected went to those support groups…what they went through before they actually got the help they needed. And, like, based off that, that should tell you that [your friend] shouldn't do* [marijuana]

Discussion

Principal Findings

Our results suggest that the KT program was acceptable and feasible for addressing the knowledge gap about cannabis and psychosis experienced by predominantly underage Black youth.

This small-scale pre-post design suggests the KT program can be used with Black youth to increase their knowledge of cannabis and psychosis risk among youth. The tutorial discussions suggest that the participants were satisfied with the KT program, which they found entertaining, relevant, and educational. The video games resonated with their personal experiences. Our approach was novel because it employed peer modeling—undergraduate students, youth in their own right with a degree of academic success, facilitated complex discussions with Black youth about scientific data. Further work is needed, but the facilitated tutorials may allow participants to gain further insight beyond what was presented in the game. Youth-driven collaboration may be a motivating factor in encouraging discussion and engagement. The online format worked well and provided opportunities to engage youth with a 1:1 facilitator/participant ratio.

Limitations

There were several limitations associated with this pilot project. Only 9 participants were enrolled in this study, which limits the generalizability of its results. The ethnic identity of the Black youth was not captured (Black Caribbean ethnicity versus Black African) in light of the small sample size. This flaw ignores significant differences in cannabis use among different ethnic groups in Black racialized communities [15], and these groups may respond differently to the same strategies.

Although the overall objectives for the tutorials were the same for each group, the facilitators had autonomy over the delivery of the tutorial content, so participants in different groups did not have the same tutorial content. However, the learning objectives were consistent among tutorial groups. The university students were trained as a group and conferred to ensure the tutorial content was similar between tutorial groups. The foundation of the knowledge—the video games—was consistent.

Most participants (8/9, 88%) reported using their smartphones to play video games, so creating a version for mobile phones might increase its accessibility. The reported frequency of video game play in the past year was low among participants. Approximately half of the participants (5/9, 56%) reported playing video games less than once per month in the past year. More specifically, the Ontario Student Drug Use and Health Survey found that among students in grades 7-12, 16.7% reported that they did not play video games, 22.6% played 3 times per month or less, 6.2% played once per week, 17.3% played 2-3 times per week, 13.0% played 4-5 times a week, and 24.3% played daily or almost daily [41]. Canadian youth who play video games more frequently might not be as impressed with the novelty of the SERIES and might expect graphic quality in keeping with commercial video games. The Back to Reality Series is a research prototype created with limited funding from research grants, hence the graphics and music are not on par with those typically found in commercial video games. Furthermore, our prototype could only be used with a Windows operating system. Additional minigames throughout gameplay might further increase user engagement and help with information recall. However, the entertainment value of the SERIES would be in competition with traditional psychoeducational programs about cannabis and mental health effects because it is geared for use by community agencies and mental health programs.

The design of this pilot fails to identify which component of the KT program had the most significant impact—the SERIES versus the tutorial sessions. Nevertheless, this KT program achieved a 37% increase in scores. This is the highest increase in scores among all of the demonstration projects and the only study to include tutorials. Therefore, the capacity of the SERIES to transfer knowledge appears to be augmented by the tutorial format. However, future research needs to establish whether the tutorial sessions enhance the learning provided by the games. The findings need to be replicated in a larger study to determine whether this KT program can be successfully implemented among a more extensive sample of Black and diverse youth populations.

The satisfaction questions may have introduced social desirability bias (“Did you like….?”). A Likert scale to assess satisfaction would introduce less bias than “Yes” and “No” responses to questions about video game satisfaction. Despite these shortcomings, it seems that participants enjoyed the gameplay experience overall based on their qualitative responses.

Comparisons With Prior Work

Video games have been shown to simulate real experiences that may help players practice decision-making and actions needed for their real life by using gamified conditions to augment learning [42]. The participants valued the entertainment and educational components of the KT program. Video games have great appeal and are widely used by youth with serious mental illnesses [43]. Our video games seemed to ground the tutorial’s content with concrete examples and thus provided opportunities to consolidate concepts introduced by the games. Participants were able to engage in rich, student-led discussions by referring to narratives from within the video games without resorting to personal confidential information. The video games seemed to be engaging and succeeded in capturing the participants’ interest.

The context or setting for gathering information was very relevant for this study. Peer collaboration seemed to be a motivating factor in encouraging discussion and engagement. The online format worked well and provided opportunities to engage youth on a 1:1 basis with university student facilitators. A 2018 study involving undergraduate students revealed that interactive online tutorials with student-led discussions and self-assessment quizzes were more effective for knowledge transfer than traditional textbook learning [17].

Conclusions

Effective strategies to educate youth about substance use risk have long been a challenge. This KT program adds a potential strategy to the list of online educational curricula. This study provides valuable insights into the feasibility and acceptability of this innovative KT program in disseminating cannabis-related health information to Black youth in Canada. Future research could increase the sample size and disaggregate the video game data from the tutorial sessions to understand the role played by each approach in educating Black youth. The KT program could also be studied in a larger, long-term trial to test whether this

knowledge acquisition leads to less problematic cannabis use among youth.

Acknowledgments

We would like to thank Dr David Green, Executive Director, and Alfia Ansu, Administrative Assistant, as well as all of the staff, participants, and community members supporting this research project at Free For All Foundation in Brampton, Ontario, Canada.

This study was funded by the Michael G. DeGroote Centre for Medicinal Cannabis Research and the Department of Psychiatry, McMaster University. The Back to Reality Series video games were funded by the Canadian Institute for Health Research, Mitacs, Ontario Trillium Foundation, AMS-The Phoenix Project, and CARSTAR Canada Research Innovation fund.

Conflicts of Interest

SA was involved in the development of the Back to Reality Series. All other authors declare no conflicts of interest.

Multimedia Appendix 1
Psychosis and cannabis test quiz.
[DOCX File , 23 KB - formative_v6i6e33693_app1.docx]

References

1. Callaghan RC, Sanches M, Benny C, Stockwell T, Sherk A, Kish SJ. Who consumes most of the cannabis in Canada? Profiles of cannabis consumption by quantity. Drug Alcohol Depend 2019 Dec 01;205:107587-107595. [doi: 10.1016/j.drugalcdep.2019.107587] [Medline: 31600617]
2. Hammond D, Wadsworth E, Reid JL, Burkhalter R. Prevalence and modes of cannabis use among youth in Canada, England, and the US, 2017 to 2019. Drug Alcohol Depend 2021 Feb 01;219:108505-108518. [doi: 10.1016/j.drugalcdep.2020.108505] [Medline: 33421799]
3. Leos-Toro C, Reid J, Madill C, Rynard V, Manske S, Hammond D. Cannabis in Canada-Tobacco Use in Canada: Patterns and Trends, 2017 Edition, Special Supplement. Propel Centre for Population Health Impact, editor. Waterloo, ON: University of Waterloo; 2017. URL: https://uwaterloo.ca/tobacco-use-canada/ [accessed 2022-08-09]
4. Sarvet AL, Wall MM, Keyes KM, Cerdá M, Schulenberg JE, O'Malley PM, et al. Recent rapid decrease in adolescents' perception that marijuana is harmful, but no concurrent increase in use. Drug Alcohol Depend 2018 May 01;186:68-74 [FREE Full text] [doi: 10.1016/j.drugalcdep.2017.12.041] [Medline: 29550624]
5. Hill K. Marijuana: The Unbiased Truth about the World's Most Popular Weed. City Centre, Minnesota: Hazelden Publishing; 2015:33-40.
6. Arseneault L, Cannon M, Witton J, Murray RM. Causal association between cannabis and psychosis: examination of the evidence. Br J Psychiatry 2004 Feb;184:110-117. [doi: 10.1192/bjp.184.2.110] [Medline: 14754822]
7. van Os J, Bak M, Hanssen R, Bijl RV, de Graaf R, Verdoux H. Cannabis use and psychosis: a longitudinal population-based study. Am J Epidemiol 2002 Aug 15;156(4):319-327. [doi: 10.1093/aje/kwf043] [Medline: 12181101]
8. Di Forti M, Sallis H, Allegri F, Trotta A, Ferraro L, Stilo SA, et al. Daily use, especially of high-potency cannabis, drives the earlier onset of psychosis in cannabis users. Schizophr Bull 2014 Nov;40(6):1509-1517 [FREE Full text] [doi: 10.1093/schbul/sbt181] [Medline: 24345517]
9. Caspi A, Moffitt TE, Cannon M, McClay J, Murray R, Harrington H, et al. Moderation of the effect of adolescent-onset cannabis use on adult psychosis by a functional polymorphism in the catechol-O-methyltransferase gene: longitudinal evidence of a gene X environment interaction. Biol Psychiatry 2005 May 15;57(10):1117-1127. [doi: 10.1016/j.biopsych.2005.01.026] [Medline: 15866551]
10. Vaucher J, Keating BJ, Lasserre AM, Gan W, Lyall DM, Ward J, et al. Cannabis use and risk of schizophrenia: a Mendelian randomization study. Mol Psychiatry 2018 May;23(5):1287-1292 [FREE Full text] [doi: 10.1038/mp.2016.252] [Medline: 28115737]
11. Di Forti M, Quattrone D, Freeman TP, Tripoli G, Gayer-Anderson C, Quigley H, et al. The contribution of cannabis use to variation in the incidence of psychotic disorder across Europe (EU-GEI): a multicentre case-control study. The Lancet Psychiatry 2019 May;6(5):427-436 [FREE Full text] [doi: 10.1016/S2215-0366(19)30048-3] [Medline: 30902669]
12. Di Forti M, Marconi A, Carra E, Fraietta S, Trotta A, Bonomo M, et al. Proportion of patients in south London with first-episode psychosis attributable to use of high potency cannabis: a case-control study. Lancet Psychiatry 2015 Mar;2(3):233-238 [FREE Full text] [doi: 10.1016/S2215-0366(14)00117-5] [Medline: 26359901]
13. Fischer B, Russell C, Sabioni P, van den Brink W, Le Foll B, Hall W, et al. Lower-Risk Cannabis Use Guidelines: A Comprehensive Update of Evidence and Recommendations. Am J Public Health 2017 Aug;107(8):e1-e12. [doi: 10.2105/AJPH.2017.303818] [Medline: 28644037]
14. Owusu-Bempah A. Amplifying Black experiences in cannabis and mental health research: Virtual dialogue series. Ottawa, Canada: National Library of Canada 2021:1-16. [doi: 10.4324/9781315686400-4]

15. Tuck A, Hamilton HA, Agic B, Ialomiteanu AR, Mann RE. Past year cannabis use and problematic cannabis use among adults by ethnicity in Ontario. Drug Alcohol Depend 2017 Oct 01;179:93-99. [doi: 10.1016/j.drugalcdep.2017.06.022] [Medline: 28763781]
16. Brody GH, Murry VM, Gerrard M, Gibbons FX, McNair L, Brown AC, et al. The strong African American families program: prevention of youths' high-risk behavior and a test of a model of change. J Fam Psychol 2006 Mar;20(1):1-11. [doi: 10.1037/0893-3200.20.1.1] [Medline: 16569084]
17. Kramer M, Olson D, Walker JD. Design and Assessment of Online, Interactive Tutorials That Teach Science Process Skills. CBE Life Sci Educ 2018 Jun;17(2):ar19 [FREE Full text] [doi: 10.1187/cbe.17-06-0109] [Medline: 29749846]
18. Clark G, Russell J, Enyeart P, Gracia B, Wessel A, Jarmoskaite I, et al. Science Educational Outreach Programs That Benefit Students and Scientists. PLoS Biol 2016 Feb 4;14(2):e1002368-e1002376 [FREE Full text] [doi: 10.1371/journal.pbio.1002368] [Medline: 26844991]
19. Wakefield MA, Loken B, Hornik RC. Use of mass media campaigns to change health behaviour. Lancet 2010 Oct 9;376(9748):1261-1271 [FREE Full text] [doi: 10.1016/S0140-6736(10)60809-4] [Medline: 20933263]
20. Ferrari M, McIlwaine SV, Reynolds JA, Archie S, Boydell K, Lal S, et al. Digital Game Interventions for Youth Mental Health Services (Gaming My Way to Recovery): Protocol for a Scoping Review. JMIR Res Protoc 2020 Jun 24;9(6):e13834-e13844 [FREE Full text] [doi: 10.2196/13834] [Medline: 32579117]
21. Baranowski T, Buday R, Thompson DI, Baranowski J. Playing for real: video games and stories for health-related behavior change. Am J Prev Med 2008 Jan;34(1):74-82 [FREE Full text] [doi: 10.1016/j.amepre.2007.09.027] [Medline: 18083454]
22. Lau HM, Smit JH, Fleming TM, Riper H. Serious Games for Mental Health: Are They Accessible, Feasible, and Effective? A Systematic Review and Meta-analysis. Front Psychiatry 2016;7:209-222 [FREE Full text] [doi: 10.3389/fpsyt.2016.00209] [Medline: 28149281]
23. Fleming TM, Bavin L, Stasiak K, Hermansson-Webb E, Merry SN, Cheek C, et al. Serious Games and Gamification for Mental Health: Current Status and Promising Directions. Front Psychiatry 2016;7:215-222 [FREE Full text] [doi: 10.3389/fpsyt.2016.00215] [Medline: 28119636]
24. Rizzo AS, Difede J, Rothbaum BO, Reger G, Spitalnick J, Cukor J, et al. Development and early evaluation of the Virtual Iraq/Afghanistan exposure therapy system for combat-related PTSD. Ann N Y Acad Sci 2010 Oct;1208:114-125. [doi: 10.1111/j.1749-6632.2010.05755.x] [Medline: 20955333]
25. Fleming T, Dixon R, Frampton C, Merry S. A pragmatic randomized controlled trial of computerized CBT (SPARX) for symptoms of depression among adolescents excluded from mainstream education. Behav Cogn Psychother 2012 Oct;40(5):529-541. [doi: 10.1017/S1352465811000695] [Medline: 22137185]
26. Merry SN, Stasiak K, Shepherd M, Frampton C, Fleming T, Lucassen MFG. The effectiveness of SPARX, a computerised self help intervention for adolescents seeking help for depression: randomised controlled non-inferiority trial. BMJ 2012 Apr 18;344:e2598 [FREE Full text] [doi: 10.1136/bmj.e2598] [Medline: 22517917]
27. Pot-Kolder RMCA, Geraets CNW, Veling W, van Beilen M, Staring ABP, Gijsman HJ, et al. Virtual-reality-based cognitive behavioural therapy versus waiting list control for paranoid ideation and social avoidance in patients with psychotic disorders: a single-blind randomised controlled trial. Lancet Psychiatry 2018 Dec;5(3):217-226. [doi: 10.1016/S2215-0366(18)30053-1] [Medline: 29429948]
28. Jenkins EK, Slemon A, Haines-Saah RJ. Developing harm reduction in the context of youth substance use: insights from a multi-site qualitative analysis of young people's harm minimization strategies. Harm Reduct J 2017 Jul 31;14(1):53-64 [FREE Full text] [doi: 10.1186/s12954-017-0180-z] [Medline: 28760146]
29. Archie S, Gyomorey K. First Episode Psychosis, Substance Abuse and Prognosis: A Systematic Review. CPSR 2009 Aug 01;5(3):153-163. [doi: 10.2174/157340009788971074]
30. Archie S, Boydell KM, Stasiulis E, Volpe T, Gladstone BM. Reflections of young people who have had a first episode of psychosis: what attracted them to use alcohol and illicit drugs? Early Interv Psychiatry 2013 May;7(2):193-199. [doi: 10.1111/j.1751-7893.2012.00355.x] [Medline: 22404861]
31. Archie S, Akhtar-Danesh N, Norman R, Malla A, Roy P, Zipursky RB. Ethnic diversity and pathways to care for a first episode of psychosis in Ontario. Schizophr Bull 2010 Jul;36(4):688-701 [FREE Full text] [doi: 10.1093/schbul/sbn137] [Medline: 18987101]
32. Ferrari M, Flora N, Anderson KK, Tuck A, Archie S, Kidd S, et al. The African, Caribbean and European (ACE) Pathways to Care study: a qualitative exploration of similarities and differences between African-origin, Caribbean-origin and European-origin groups in pathways to care for psychosis. BMJ Open 2015 Jan 14;5(1):e006562-ee00674 [FREE Full text] [doi: 10.1136/bmjopen-2014-006562] [Medline: 25588783]
33. Winn B. The design, play, experience framework. In: Handbook of research on effective electronic gaming in education. Hershey, PA: IGI Global; 2009:1011-1021.
34. Ferrari M, Suzanne A. Youth Mental Health, Family Practice, and Knowledge Translation Video Games about Psychosis: Family Physicians' Perspectives. J Can Acad Child Adolesc Psychiatry 2017;26(3):184-189 [FREE Full text] [Medline: 29056980]
35. Chiu J. Back to Reality Series Psychoeducational Video Games for Increasing Knowledge of Psychosis associates with Marijuana Use. Psynapse: McMaster Undergraduate Journal of Psychology, Neuroscience, and Behaviour 2021;4(5):55-56.

36. Archie S, Ferrari M, Bzonek P. Back to Reality Series Video Games; Translating Research Knowledge about Marijuana and Psychosis for Youth. 2018 Oct 10 Presented at: IEPA 11th International Conference on Early Intervention in Mental Health; October 10, 2018; Boston, Massachusetts, USA. [doi: 10.1111/eip.12722]
37. Douglas A, Archie S. Use of video game technology as an early intervention tool for youth experiencing psychosis. 2015 Presented at: 168th American Psychiatric Association Annual Meeting; May 16, 2015; Toronto, ON p. 2015.
38. Archie S, Ferrari M, Naidu A, Douglas A. Back to Reality Series: Using arts-based research to create video games about psychosis and pathways to care. 2016 Presented at: IEPA 10th International Conference on Early Intervention in Mental Health - "Looking Back, Moving Forward"; October 22, 2016; Milan, Italy.
39. Dasgupta P, Tureski T, Lenzi R, Bindu K, Nand G. Half the Sky Movement Multimedia Communication Initiative: An Evaluation of the 9-Minutes Mobile Game and Video. C-Change: Innovative approaches to social and behavior change communication. 2012. URL: https://c-changeprogram.org [accessed 2015-07-18]
40. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
41. Boak A, Hamilton H, Adlaf E, Henderson J, Mann R. The mental health and well-being of Ontario students, 1991-2019. In: Detailed from the Ontario Student Drug Use and Health Survey. Toronto: Centre for Addiction and Mental Health, 2016; 2019:1-244.
42. Gee JP. Learning by Design: Good Video Games as Learning Machines. E-Learning and Digital Media 2016 Jul 31;2(1):5-16. [doi: 10.2304/elea.2005.2.1.5]
43. Abdel-Baki A, Lal S, D-Charron O, Stip E, Kara N. Understanding access and use of technology among youth with first-episode psychosis to inform the development of technology-enabled therapeutic interventions. Early Interv Psychiatry 2015 May 22:72-76. [doi: 10.1111/eip.12250] [Medline: 26011657]

Abbreviations

CBT: cognitive behavioral therapy
KT: knowledge translation
OR: odds ratio
PCT: psychosis and cannabis test
THC: delta-9-tetrahydrocannabinol

Edited by A Mavragani; submitted 23.09.21; peer-reviewed by J Fleming; comments to author 23.12.21; revised version received 29.01.22; accepted 17.03.22; published 20.06.22.

Please cite as:
Jani P, Song N, Artna E, Lyeo J, Assam A, Maelzer F, Murphy A, Grant A, Archie S
Online Knowledge Translation Program Involving Video Games and University Student–Led Tutorials About Cannabis and Psychosis for Black Youth: Mixed Method Feasibility Study
JMIR Form Res 2022;6(6):e33693
URL: https://formative.jmir.org/2022/6/e33693
doi:10.2196/33693
PMID:35315782

Original Paper

A Telehealth-Delivered Tai Chi Intervention (TaiChi4Joint) for Managing Aromatase Inhibitor–Induced Arthralgia in Patients With Breast Cancer During COVID-19: Longitudinal Pilot Study

Sameh Gomaa[1], MD; Carly West[1], MPH; Ana Maria Lopez[1], MD; Tingting Zhan[2], PhD; Max Schnoll[1], BA; Maysa Abu-Khalaf[1], MD; Andrew Newberg[3], MD; Kuang-Yi Wen[1], PhD

[1]Department of Medical Oncology, Thomas Jefferson University, Philadelphia, PA, United States

[2]Department of Pharmacology & Experimental Therapeutics, Thomas Jefferson University, Philadelphia, PA, United States

[3]Department of Integrative Medicine and Nutritional Sciences, Thomas Jefferson University, Philadelphia, PA, United States

Corresponding Author:
Kuang-Yi Wen, PhD
Department of Medical Oncology
Thomas Jefferson University
834 Chestnut Street Suite 300
Philadelphia, PA, 19106
United States
Phone: 1 2155034623
Email: Kuang-Yi.Wen@jefferson.edu

Related Article:

This is a corrected version. See correction statement: https://formative.jmir.org/2022/7/e40830

Abstract

Background: Estrogen receptor–positive breast cancer is the most common type of breast cancer in postmenopausal women. Aromatase inhibitors (AIs) are the endocrine therapy of choice recommended for these patients. Up to 50% of those treated with an AI develop arthralgia, often resulting in poor adherence and decreased quality of life.

Objective: The study is a single-arm longitudinal pilot study aiming to evaluate the safety, feasibility, acceptability, and potential efficacy of *TaiChi4Joint*, a remotely delivered 12-week tai chi intervention designed to relieve AI-induced joint pain.

Methods: Women diagnosed with stage 0-III breast cancer who received an AI for at least 2 months and reported arthralgia with a ≥4 score on a 0 to 10 scale for joint pain were eligible for study enrollment. Participants were encouraged to join tai chi classes delivered over Zoom three times a week for 12 weeks. Program engagement strategies included using a private Facebook study group and a Box cloud for archiving live class recordings. The program uses SMS text messaging and emails with periodic positive quotes and evidence-based information on tai chi for facilitating community bonding and class attendance. Participants were invited to complete the following assessments at baseline and at 1-, 2-, and 3-month intervals from study enrollment: Brief Pain Inventory, Western Ontario and McMaster University Osteoarthritis Index (WOMAC), The Australian Canadian Osteoarthritis Hand Index (AUSCAN), Fatigue Symptom Inventory, Hot Flash Related Daily Interference Scale (HFRDIS), Pittsburgh Sleep Quality Index (PSQI), and Center for Epidemiological Studies–Depression (CES-D).

Results: A total of 55 eligible patients were invited to participate, and 39 (71%) consented and completed the baseline assessments. Participants attended 61% (median) of the suggested classes, with no tai chi–related adverse events reported. Of the 39 participants, 22 completed the 3-month follow-up assessment with a 56% retention rate. Study participants reported improvement from baseline compared to 3 months as follows (paired *t* test): Brief Pain Inventory (P<.001), AUSCAN pain subscale (P=.007), AUSCAN function subscale (P=.004), Fatigue Symptom Inventory (P=.004) and PSQI (P<.001), and HFRDIS (P=.02) and CES-D (P<.001). In particular, for our primary end point of interest, improvements in hip and knee symptoms, measured by WOMAC's three subscales, were clinically meaningful and statistically significant when adjusted for multiple comparisons from baseline to 3 months post intervention.

Conclusions: The COVID-19 global pandemic has resulted in the need to rethink how mind-body therapies can be delivered. This study demonstrated the feasibility, acceptability, and potential efficacy of a telehealth-based tai chi intervention for reducing AI-induced arthralgia. The intervention decreased patient-reported pain and stiffness, and improved sleep quality and depressive symptoms. Fully powered, large, telehealth-based tai chi trials for AI-associated arthralgia are needed considering our promising findings.

Trial Registration: ClinicalTrials.gov NCT04716920; https://www.clinicaltrials.gov/ct2/show/NCT04716920

(JMIR Form Res 2022;6(6):e34995) doi:10.2196/34995

KEYWORDS

breast cancer; arthralgia; tai chi; telehealth, pain; mind-body therapy

Introduction

Breast cancer (BC) is the most common cancer among women in the United States, and hormone receptor–positive BC accounts for approximately two-thirds of all BC. For postmenopausal women with hormone receptor–positive BC [1], long-term use of an aromatase inhibitor (AI) for 5 to 10 years after primary surgical treatment has been demonstrated to prevent disease relapse and improve disease-free survival [2-4]. A meta-analysis revealed that the overall pooled prevalence rate of AI-induced arthralgia was 45.9% in postmenopausal hormone receptor–positive patients with BC [5]. AI-related arthralgia includes the new onset of worsening joint pain, stiffness, and impaired function. Arthralgia can affect any joints but primarily affects the hips, knees, and wrists/hands, with 70% of women describing their joint pain as moderate to severe [6]. Symptom onset is generally within 6 weeks to 12 months after AI initiation but can occur even longer, and symptoms may appear abruptly or increase gradually over time [6-8]. The clinical significance of AI-related arthralgia is that in addition to negatively affecting the quality of life [9] and daily activities [10], it can also decrease patients' adherence to AI use. Arthralgia can lead up to 25% of patients with BC who stop AIs early and 50% of patients who have disruptions to their treatment regimen and schedule [6,10]. This data has important clinical implications, as nonadherence to AIs in early BC has been shown to negatively impact BC-related survival [10]. Other common symptoms during AI therapy include fatigue [11], hot flashes [12], insomnia [13], and depression [14,15], all of which can also adversely affect a patient's quality of life and AI treatment adherence [15].

The approach to managing AI-related arthralgia is complex due to the absence of validated treatment standards for clinicians to use. They are complicated by limited knowledge regarding the physiological mechanisms involved. Unfortunately, pharmacological interventions to treat AI-related arthralgia have resulted in limited relief [16] and can produce undesirable side effects and physiologic dependence, especially in elderly patients. Lifestyle changes and complementary and alternative interventions, including supplementation, physical activities, and acupuncture, have been evaluated in a small number of studies with mixed results [16]. Thus, there is a growing interest in developing and assessing novel integrative programs to manage this adverse event and improve patient adherence to AI therapy.

Tai chi is a multidimensional mind-body therapy that integrates moderate physical activity, deep breathing, and meditation; additionally, it offers a promising approach to symptom management in cancer populations [17]. Tai chi has been practiced for centuries, and approximately 25% of the urban female population in the United States older than 50 years has practiced tai chi [18]. Tai chi forms consist of a series of upper- and lower-extremity movements performed in a particular choreographic manner. The interplay modes may provide a possible rationale for soft and connective tissues [17]. These movements may induce local biochemical changes that modulate blood circulation, improve muscle flexibility, intensify the action of the lymphatic system, and loosen adherent connective tissue. The loosening of the adherent connective tissue is thought to enhance the reuptake of local nociceptive and inflammatory mediators [17]. Randomized clinical trials (RCTs) have demonstrated the feasibility and efficacy of tai chi for women with BC. Studies have reported tai chi's beneficial effects on psychological status (eg, depression or self-efficacy), physical function (eg, upper limb functional mobility), and other symptoms (eg, fatigue or sleep dysfunction) [19].

Furthermore, tai chi interventions designed specifically for individuals with pain-related conditions (eg, osteoarthritic knee pain) have significantly improved pain and physical functional outcomes [20]. However, there is lack of data regarding the use of tai chi interventions to target Al-induced arthralgia for patients with BC. Almost no tai chi interventions exist using a remote telehealth administration for any contexts. Telehealth or telemedicine uses telecommunications technologies for increased access to health care. Several advantages exist for telehealth, including the cost- and time-effectiveness, especially amid the current COVID-19 pandemic [21,22], as cancer survivors are at a higher risk of developing complications from COVID-19 [23,24]. Their risk of contracting COVID-19 must be reduced. New information and communication technology (eg, Zoom and WhatsApp) offers a convenient solution. Aligned with the social distancing and quarantine requirements, between December 2020 to July 2021, we conducted a pilot study that evaluated the feasibility, acceptability, and preliminary efficacy of a 12-week tai chi intervention (*TaiChi4Joint*). The study was aimed to reduce AI-induced arthralgia in women with BC, delivered remotely over the Zoom videoconferencing platform (Zoom Video Communications) facilitated by social media and SMS text messaging–enabled engagement strategies.

Methods

Study Eligibility and Study Procedures

Inclusion criteria for study participation included participants being ≥18 years of age, being able to speak/read English, having been diagnosed with stage 0-III BC, being postmenopausal (ie, no menses for at least 1 year), receiving an AI (anastrozole, letrozole, or exemestane) for at least 2 months, having joint pain that started or worsened after the initiation of AIs, reporting that their worst pain score in the prior week was ≥4 score on a 0 to 10 scale, and willing to adhere to all the study procedures.

Exclusion criteria included having another type of cancer that was diagnosed in the past 5 years; having uncontrolled cardiac, pulmonary, or infectious disease; having a BMI>40 kg/m^2; currently attending any mind-body therapy classes (eg, tai chi or yoga); having joint pain due to an inflammatory arthritic condition; having surgery in the past 6 months; having a joint injection in the past 3 months; currently taking corticosteroids or opioids; or having discontinued or planning to discontinue their AI in the next 6 months.

A research team member screened potentially eligible patients via the electronic medical record or were referred by the patient's treating medical oncologists. The study team connected with major support groups in the greater Philadelphia area to promote the study and potentially reach eligible participants. An informed consent form was signed, mailed, and returned by a study participant after reviewing the study rationale, intervention, and potential adverse effects with a research team member. The research team also provided remote assistance over the phone to ensure that the Zoom platform was successfully installed on the participant's computer, tablet, or smartphone and that they could access classes on time. Participants were instructed to complete baseline, 1-month, 2-month, and 3-month surveys either by a mailed paper-based survey or via a REDCap online survey.

Ethical Approval

This study with IRB control number 20G.093 was administratively approved by Thomas Jefferson University Institutional Review Board on January 30, 2020 by Board number 2405. Study accrual began in December 2020 and was completed in April 2021.

TaiChi4Joint Study Intervention

Tai chi is a multicomponent practice that integrates physical, psychosocial, emotional, spiritual, and behavioral elements. Participants were invited to attend three tai chi classes per week for the 12-week study duration via the Zoom platform. Using a manualized approach, each class provided objectives and learning activities. The learning activities included sequentially learning a specific set of tai chi that included 24 movements with verification of skills attainment on a weekly basis. Each class began with relaxation exercises that had breathing exercises and qigong warm-ups. The instructor reviewed the previously learned techniques and introduced new movements following the relaxation session. The first 8 weeks of classes focused on mastery of single forms through multiple repetitions. The latter weeks emphasized consolidation of daily practice routines with natural breathing integrated into all classes. Scheduled classes ran from 45 to 60 minutes each and occurred 4 times per week, thus allowing patients to choose 3 classes that best fit their schedule. Participants were also encouraged to practice these tai chi classes through video recordings and spend 30 minutes daily on most days of the week self-practicing the technique they learned.

Facebook and Text Messaging–Enabled Engagement Strategies

Participants were encouraged to join an optional Facebook private *TaiChi4joint* group that consists of instructional videos matching the progress of weekly classes for at-home practice and promotes peer support in tai chi engagement. We used the Facebook page and a Box shared cloud drive to share the instructional video and the recorded live tai chi sessions for participants who could not attend class or wanted additional practice. Weekly Facebook posts also contained evidence-based information about tai chi, the potential impact of tai chi in reducing pain, class reminders, and milestone celebrations (eg, birthday or program completion) to foster participant engagement and community building. Further, weekly text messages addressing class schedule, self-practice reminders, and brief positive quotes were proactively sent to participants to promote engagement and offer additional support.

Outcome Measures

Patient Demographics and Treatment Information

We collected the following study subject demographics and BC-specific information: age, race, ethnicity, education, marital status, BC diagnosis and stage, treatment history, and AI medication type.

Feasibility

We aimed to retain 50% of eligible patients at the 3-month follow-up survey. The cutoff of 50% retention rate was chosen and derived from other similar pilot studies using mind-body therapies to mitigate arthralgia [24]. Feasibility was also measured by the participant's class attendance, which the research team coordinator logged each session.

Acceptability

The acceptability of the study intervention was defined as a 60% consenting rate. Justification for a 60% consenting rate as acceptable was based on prior BC survivorship behavioral studies [25,26]. We also asked whether participants found the intervention helpful in reducing their pain.

Musculoskeletal Symptoms

As arthralgia primarily affects the hips and knees [6], we chose the Western Ontario and McMaster University Osteoarthritis Index (WOMAC) as the primary efficacy outcome. WOMAC measures lower-extremity joint symptoms (hips/knees) in the past 7 days in three domains: pain, stiffness, and physical function [27]. WOMAC has been recommended with great sensitivity for measuring changes in musculoskeletal symptoms in patients with BC receiving AIs [28].

Hand Pain and Physical Function

We used the Australian Canadian Osteoarthritis Hand Index (AUSCAN) to assess pain and physical function in the hands. AUSCAN has been reported to have excellent sensitivity and responsiveness in detecting and measuring AI-induced arthralgia [28].

Overall Pain

The Brief Pain Inventory (BPI) is a 14-item questionnaire developed for use in patients with cancer that assesses the worst pain, pain severity, and pain interference over the past week reported on a scale of 0 to 10. The BPI is the most common, valid, and reliable measure to assess pain in patients with cancer.

Fatigue

The Fatigue Symptom Inventory (FSI) measure is a 13-item self-report measure shown to be sensitive to assessing change in fatigue among patients with BC and with substantial internal consistency. An overall score was used.

Hot Flash

We used the Hot Flash Related Daily Interference Scale (HFRDIS) to measure the effect of hot flashes on the overall quality of life and nine specific activities: work, social activities, leisure activities, sleep, mood, concentration, relations with others, sexuality, and enjoyment of life [29].

Sleep Quality

We used the Pittsburgh Sleep Quality Index (PSQI) for a subjective sleep assessment, including multiple sleep-related variables over the preceding month [30].

Depressive Symptoms

We used the Center for Epidemiological Studies–Depression (CES-D) to assess depressive symptoms [31], and it has shown excellent internal consistency and validity in patients with cancer [32].

Analysis

The statistical analysis was performed using R (R Foundation for Statistical Computing). Summary demographic statistics were calculated. Summary statistics of the end points, WOMAC subscales, AUSCAN subscales, BPI, FSI, HFRDIS, PSQI, and CES-D at baseline and at 1-month, 2-month, and 3-month follow-ups were calculated. Paired *t* tests of 3 months versus baseline were performed for each end point, and findings with 95% CIs are presented. For our primary end points of the WOMAC subscales, clinically meaningful changes in the subscales were evaluated using the criteria >1.5 out of 20 decrease for WOMAC pain, >0.6 out of 8 decrease for WOMAC stiffness, and >4.6 out of 68 decrease for WOMAC function score [32]. In addition, the linear time trends of end points were explored using latent class mixed-effect models with the R package lcmm. Up to 3 latent classes were considered, and the best number of latent classes was determined by the Bayesian information criterion.

Results

Patient Characteristics

The mean age of patients was 58 (SD 10.6) years. Of the 39 patients, 30 were Caucasian and 36 were at least college graduates. There were 31 patients with stage I or II, and had undergone mastectomy (n=22), radiation (n=27), or chemotherapy (n=23). Half of the participants (n=21) had been taking anastrozole. A total of 24 patients had been taking an AI for less than 3 years (Table 1).

Table 1. Participant characteristics.

Variable	Participants (N=39)
Age (years), mean (SD)	58.18 (10.60)
Race, n (%)	
Caucasian	30 (77)
African American	7 (18)
Asian	1 (3)
Unreported	1 (3)
Ethnicity, n (%)	
Non-Hispanic	38 (97)
Hispanic	1 (3)
Marital status, n (%)	
Married	26 (67)
Single	7 (18)
Divorced	5 (13)
Widowed	1 (3)
Education level, n (%)	
High school	3 (8)
College graduate	16 (41)
Postgraduate school	20 (51)
Breast cancer stage, n (%)	
0	2 (5)
I	14 (36)
II	17 (44)
III	6 (15)
Surgery[a], n (%)	
Mastectomy	22 (56)
Lumpectomy	15 (39)
Radiation[a], n (%)	27 (69)
Chemotherapy[a], n (%)	23 (59)
AI[b] type, n (%)	
Letrozole	6 (15)
Anastrozole	21 (54)
Exemestane	9 (23)
Other	3 (8)
AI medication duration (years), n (%)	
≤1	12 (31)
1-3	12 (31)
3-6	12 (31)
Unreported	3 (8)

[a]Treatment options do not add up to 39 due to individual treatment choices.

[b]AI: aromatase inhibitor.

Acceptability

A total of 55 eligible patients were invited to participate, and 39 of the 55 (70.9%, 95% exact CI 57.1%-82.4%) patients consented to study participation and completed the baseline assessment, which exceeded the acceptability threshold of 60%. To seek participants' feedback on the intervention, 14 of 22 participants who completed the final survey responded to our postintervention evaluation. All 14 perceived high satisfaction with the intervention, and 10 reported that the intervention was beneficial in relieving their pain.

Feasibility

Of the 39 participants, 26 completed a 1-month follow-up, 21 completed a 2-month follow-up, and 22 completed the final 3-month follow-up, with a retention rate of 56% (95% exact CI 39.6%-72.2%) at the end of the study. Study attendance was calculated as the number of classes of the 36 suggested classes (12 weeks × 3 per week) each participant attended, ranging from 3% (2 classes) to 100% (36 classes), with a median value of 61% (22 classes). Of the 39 participants, 16 attended 18 (50%) or more classes during the 12-week study period.

Change in Patient-Reported Outcomes During the Study Duration

Table 2 shows statistically significant improvement in all our study measures from baseline to the 3-month follow-up. Overall pain (BPI) was reduced from 5.04 to 2.69 (P<.001). Significant reductions also occurred in the pain subscale (from 6.64 to 3.52; P=.007) and function subscale (from 11.13 to 5.71; P=.004) of AUSCAN. Significant improvements were detected in fatigue (FSI from 53.40 to 29.74; P=.004), hot flashes (HFRDIS from 28.14 to 12.44; P=.02), sleep quality (PSQI from 18 to 11.1; P<.001), and depressive symptom (CES-D from 38.33 to 33.95; P=.03). In particular, the significant reductions in pain (from 8.05 to 5.05; P<.001), stiffness (from 4.36 to 2.33; P<.001), and function (from 23.28 to 11.67; P<.001) measured by WOMAC, detected using paired t tests, were clinically meaningful. The three P values for the WOMAC measurements were adjusted for multiple comparison using the Benjamini-Hochberg correction [33]. We also provide the linear time trends estimated using latent class mixed-effect models, with the three WOMAC measurements adjusted for multiple comparison, as referenced in Table 2.

Table 2. Change in patient-reported outcomes over time.

Measure (range)	Baseline (N=39), mean (SD)	1 month (n=25), mean (SD)	2 months (n=22), mean (SD)	3 months (n=21), mean (SD)	3-month change (95% CI)[a]	*P* value	Linear time trend detected by the latent class model	*P* value
WOMAC[b] pain (0-20 maximal pain)	8.05 (4.19)	6.52 (3.70)	6.36 (4.82)	5.05 (3.89)	2.57 (1.27-3.87)	<.001[c]	–0.51	<.001[c]
WOMAC stiffness (0-8 maximal stiffness)	4.36 (1.74)	3.44 (1.61)	3.00 (1.75)	2.33 (1.65)	1.67 (0.81-2.52)	<.001[c]	–0.77	<.001[c]
WOMAC function (0-68 minimal function)	23.28 (14.21)	17.76 (12.36)	16.32 (13.81)	11.67 (12.70)	8.95 (5.65-12.25)	<.001[c]	–0.71	<.001[c]
BPI[d] (0-10 worse pain)	5.04 (2.17)	4.22 (2.29)	4.16 (1.82)	2.69 (2.18)	2.25 (1.27-3.24)	<.001	–0.456	<.001
AUSCAN[e] pain (0-50 worse pain)	6.64 (4.57)	6.00 (4.27)	4.77 (4.08)	3.52 (3.56)	2.33 (0.71-3.96)	.007	–0.642	<.001
AUSCAN function (0-90 minimal function)	11.13 (7.63)	8.80 (8.10)	7.45 (7.97)	5.71 (7.27)	4.48 (1.62-7.33)	.004	–0.794	<.001
FSI[f] (0-130 worse fatigue)	53.40 (27.29)	40.04 (26.53)	36.94 (22.02)	29.74 (20.05)	17.53 (6.20-28.86)	.004	–0.61	<.001
HFRDIS[g] (0-100 worse hot flash)	28.14 (25.46)	19.27 (22.89)	13.10 (14.25)	12.44 (13.78)	10.15 (2.04-18.26)	.02	For 28 patients: –0.176; for 11 patients: –1.235	For 28 patients: .14; for 11 patients: <.001
PSQI[h] (0-21 lower sleep quality)	18.00 (5.86)	15.04 (6.52)	14.55 (5.70)	11.10 (6.40)	6.14 (3.32-8.96)	<.001	–0.626	<.001
CES-D[i] (0-60 greater depressive symptom severity)	38.33 (5.61)	37.40 (4.53)	36.05 (4.46)	33.95 (4.97)	2.62 (0.26-4.97)	.03	–0.353	<.001

[a]Paired *t* test.

[b]WOMAC: Western Ontario and McMaster University Osteoarthritis Index.

[c]Adjusted for multiple comparison.

[d]BPI: Brief Pain Inventory.

[e]AUSCAN: Australian Canadian Osteoarthritis Hand Index.

[f]FSI: Fatigue Symptom Inventory.

[g]HFRDIS: Hot Flash Related Daily Interference Scale.

[h]PSQI: Pittsburgh Sleep Quality Index.

[i]CES-D: Center for Epidemiological Studies-Depression.

Participant Feedback and Lessons Learned

Comments and suggestions from participants, the tai chi instructor, and our research team's observation notes were qualitatively summarized and discussed to reach a consensus on each identified theme and areas for future intervention improvement.

Improved Pain and Stiffness

Participants reported that their pain or stiffness improved, with the change starting quickly after the first few weeks of the intervention. The reduction in pain also facilitated other improvements in daily activities.

> *Stiffness and pain decreased. I felt joints in knees and elbows move into my sockets at some point. I'm more limber and able to lift my knees higher to run faster and further.*
>
> *Within 2 weeks, I had relief from the joint pain, especially my shoulder pain.*

Better Relaxation and Balance

Breathing training and exercise were recognized as helpful in improving relaxation and calmness. Many participants also reported their balance was improved over time.

> *I particularly liked the breathing exercise at the beginning. I will continue to use it for relaxation and mindfulness.*
>
> *My flexibility and balance have improved.*

Value in Instructor Support and Facebook Group/Text Messaging Motivation

Our tai chi instructor was perceived as highly supportive and patient. A bonding relationship between the instructor and participants was observed. Archived class videos and motivational words and encouragements delivered in the Facebook private group and SMS text messaging were also reported as positive and helpful, motivating participant's tai chi practice.

> *The instructor was amazing and showed extraordinary patience.*
>
> *It was helpful to have all the classes on the FB page.*
>
> *I enjoyed the occasional random words of kindness.*

The Convenience of Virtual Classes

Participants commented on the ease and comfort of virtually attending our tai chi classes, overcoming in-person participation barriers such as travel burdens and demanding schedules.

> *The Zoom class made it easier to fit into my schedule.*
>
> *I didn't have to deal with parking.*

The Limited Class Schedule

The main reason for not being able to attend scheduled classes was the participant's conflicting schedules. Participants reported that their nonattendance generally reflected competing life demands rather than their lack of desire to attend classes. Due to the budget constraint of the pilot, the availability of our weekly offered courses was limited with a fixed schedule of four classes per week.

> *Would like to have more time options during the week for live classes.*
>
> *Unfortunately, I have only been able to attend classes twice per week, and several times my schedule prevented me from one of them.*

Challenges of Varying Skill Levels in the Same Class

For the pilot, we used rolling enrollment, so participants started the course as they enrolled. However, participants and the instructor observed and reported issues of participants with different levels of tai chi learning progression, reducing the efficiency of the class conduct.

> *There were beginners starting regularly, so the pace had to remain slow; it was difficult for me to move very slowly when I knew the next move.*

The Initial Technological Difficulty, Bandwidth Limitation, and Online Streaming Challenges

We observed a learning curve for using the Zoom technology when participants initially started the intervention. Some of our older participants who joined our Zoom classes using their cellphones experienced viewing difficulties due to the small screen, their vision limitations, and interrupted internet signal. The problem of following the instructor's side-to-side movements virtually was noted.

> *Having the class on Zoom presented some challenges initially.*
>
> *It would have been helpful to have a third camera; it was difficult keeping perspectives –rights and lefts.*

Discussion

A Telehealth-Delivered Tai Chi Intervention Is Feasible and Acceptable

AI-induced arthralgia can prohibit normal functioning and decrease affected patients' quality of life. Clinical trial data show a 22% rate of AI discontinuation due to AI-induced arthralgia [12]. Real-world data suggest that the rate of AI-related arthralgia is as high as 50% [5], with minimal nonpharmacological options. Our findings support the feasibility, acceptability, and potential benefits of our *TaiChi4Joint* intervention and provide direction for future research. Due to COVID-19 social distancing constraints, our entire study implementation procedures were remotely conducted from the end of 2021 into early 2022. Concerning acceptability, 28 of the 39 eligible patients agreed to participate in our tai chi intervention delivered by Zoom. Many commented that the main reason they were attracted to joining was the convenience of modality, allowing them to attend classes from home during COVID-19. With regard to feasibility, 22 of the 39 participants who completed the baseline assessment completed the final assessment, and no safety issues were reported. The documented 61% average class attendance rate is not ideal, with less than 50% of participants attending half of the suggested class dosage. Many participants reported their nonattendance was generally due to competing work and family demands, conflicting with the schedule of the classes offered, as observed in the qualitative feedback. Offering a more flexible class schedule with different skill levels in future trials could also help to boost attendance rates by allowing participants to choose classes at a time more convenient for them.

Tai Chi Is Effective in Relieving AI-Induced Symptoms

We also found statistically significant improvements in pain, stiffness, impaired function, fatigue, hot flash, sleep quality, and depressive symptoms. In particular, the change of the WOMAC subscale scores over time, our primary outcome of interest, was clinically meaningful based on prior literature. Qualitative findings, including participants' perceived reduction in pain and stiffness, improvement in relaxation and balance, and the value of the instructor's and the intervention's support, could potentially have facilitated the quantitative change in outcomes observed in the study. Further, to counter potential digital divide challenges and improve participant engagement, the following strategies addressing our lessons learned might be helpful for future studies: include a user manual and training session practicing joining the videoconference, offer loaner iPads with cellular plans, use a multi-camera streaming setup, use closed session groups, and offer a more flexible class schedule. Our results demonstrate the potential of the *TaiChi4Joint* telehealth approach and support a fully powered RCT of *TaiChi4Joint* in the future.

Limitations

This is a single-arm pilot study; this design was the most suitable to tackle the question of feasibility and acceptability. Study

outcomes will need more scrutiny under an RCT design with a control group. This will be required to correct for possible selection bias or placebo effects. Selection bias might have been a cause for the study's favorable results, as patients adhering to the investigation have obtained benefits derived from tai chi practice.

A longer intervention duration and a more extended period of follow-up post intervention might be needed to understand the long-term and maintenance effects of tai chi in reducing AI-induced arthralgia.

The majority of our sample were Caucasian and had a college degree, limiting the generalization of our findings. Innovative recruitment strategies will be needed in the future to ensure the sample is generalizable and to increase outreach and support to minorities.

Conclusion

COVID-19 increased stress levels while reducing access to mind-body services in patients with cancer. Research on remote delivery of integrative, complementary, and alternative medicine health supports the feasibility and benefits of these services [22,34]. Our study used existing technologies to bring tai chi to patients' homes. We found that tai chi reduced joint pain and stiffness; decreased fatigue; improved sleep quality, hot flash, and depressive symptoms; and improved functioning. Tai chi was well tolerated, and no adverse events specific to the study intervention occurred. Our lessons learned and strategies described could help inform the design of telehealth-based behavioral support. Future randomized controlled trials are needed to establish the comparative efficacy of the TaiChi4Joint intervention to improve outcomes related to AI-induced arthralgia.

Acknowledgments

We would like to thank Mr Robert Weiss for his efforts and commitment as a tai chi instructor in our study. We would like to acknowledge Ms Lonaya Baker and Rita Smith for their assistance in recruiting patients. Thanks to Mr Michaël Bijker for providing the soundtrack for the instructional videos. This study was supported by the BUCK Cancer Foundation Medical Research Grant and the Sidney Kimmel Cancer Center Support Grant (National Cancer Institute Award 5P30CA056036-22).

Conflicts of Interest

None declared.

References

1. Breast cancer. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/cancer/breast/index.htm [accessed 2021-11-12]
2. Murphy CC, Bartholomew LK, Carpentier MY, Bluethmann SM, Vernon SW. Adherence to adjuvant hormonal therapy among breast cancer survivors in clinical practice: a systematic review. Breast Cancer Res Treat 2012 Jul;134(2):459-478 [FREE Full text] [doi: 10.1007/s10549-012-2114-5] [Medline: 22689091]
3. Crew KD, Greenlee H, Capodice J, Raptis G, Brafman L, Fuentes D, et al. Prevalence of joint symptoms in postmenopausal women taking aromatase inhibitors for early-stage breast cancer. J Clin Oncol 2007 Sep 01;25(25):3877-3883. [doi: 10.1200/JCO.2007.10.7573] [Medline: 17761973]
4. Burstein HJ, Prestrud AA, Seidenfeld J, Anderson H, Buchholz TA, Davidson NE, American Society of Clinical Oncology. American Society of Clinical Oncology clinical practice guideline: update on adjuvant endocrine therapy for women with hormone receptor-positive breast cancer. J Clin Oncol 2010 Aug 10;28(23):3784-3796 [FREE Full text] [doi: 10.1200/JCO.2009.26.3756] [Medline: 20625130]
5. Beckwée D, Leysen L, Meuwis K, Adriaenssens N. Prevalence of aromatase inhibitor-induced arthralgia in breast cancer: a systematic review and meta-analysis. Support Care Cancer 2017 May;25(5):1673-1686. [doi: 10.1007/s00520-017-3613-z] [Medline: 28204994]
6. Henry NL, Azzouz F, Desta Z, Li L, Nguyen AT, Lemler S, et al. Predictors of aromatase inhibitor discontinuation as a result of treatment-emergent symptoms in early-stage breast cancer. J Clin Oncol 2012 Mar 20;30(9):936-942 [FREE Full text] [doi: 10.1200/JCO.2011.38.0261] [Medline: 22331951]
7. Hershman DL, Shao T, Kushi LH, Buono D, Tsai WY, Fehrenbacher L, et al. Early discontinuation and non-adherence to adjuvant hormonal therapy are associated with increased mortality in women with breast cancer. Breast Cancer Res Treat 2011 Apr;126(2):529-537 [FREE Full text] [doi: 10.1007/s10549-010-1132-4] [Medline: 20803066]
8. Andrykowski MA, Schmidt JE, Salsman JM, Beacham AO, Jacobsen PB. Use of a case definition approach to identify cancer-related fatigue in women undergoing adjuvant therapy for breast cancer. J Clin Oncol 2005 Sep 20;23(27):6613-6622 [FREE Full text] [doi: 10.1200/JCO.2005.07.024] [Medline: 16170168]
9. Su HI, Sammel MD, Springer E, Freeman EW, DeMichele A, Mao JJ. Weight gain is associated with increased risk of hot flashes in breast cancer survivors on aromatase inhibitors. Breast Cancer Res Treat 2010 Nov;124(1):205-211. [doi: 10.1007/s10549-010-0802-6] [Medline: 20182796]

10. Desai K, Mao JJ, Su I, Demichele A, Li Q, Xie SX, et al. Prevalence and risk factors for insomnia among breast cancer patients on aromatase inhibitors. Support Care Cancer 2013 Jan;21(1):43-51 [FREE Full text] [doi: 10.1007/s00520-012-1490-z] [Medline: 22584732]
11. Shi Q, Giordano SH, Lu H, Saleeba AK, Malveaux D, Cleeland CS. Anastrozole-associated joint pain and other symptoms in patients with breast cancer. J Pain 2013 Mar;14(3):290-296 [FREE Full text] [doi: 10.1016/j.jpain.2012.11.010] [Medline: 23452648]
12. Stanton AL, Petrie KJ, Partridge AH. Contributors to nonadherence and nonpersistence with endocrine therapy in breast cancer survivors recruited from an online research registry. Breast Cancer Res Treat 2014 Jun;145(2):525-534. [doi: 10.1007/s10549-014-2961-3] [Medline: 24781972]
13. Yang GS, Kim HJ, Griffith KA, Zhu S, Dorsey SG, Renn CL. Interventions for the treatment of aromatase inhibitor-associated arthralgia in breast cancer survivors: a systematic review and meta-analysis. Cancer Nurs 2017;40(4):E26-E41. [doi: 10.1097/NCC.0000000000000409] [Medline: 27333128]
14. Greenlee H, Crew KD, Shao T, Kranwinkel G, Kalinsky K, Maurer M, et al. Phase II study of glucosamine with chondroitin on aromatase inhibitor-associated joint symptoms in women with breast cancer. Support Care Cancer 2013 Apr;21(4):1077-1087 [FREE Full text] [doi: 10.1007/s00520-012-1628-z] [Medline: 23111941]
15. Khan QJ, Reddy PS, Kimler BF, Sharma P, Baxa SE, O'Dea AP, et al. Effect of vitamin D supplementation on serum 25-hydroxy vitamin D levels, joint pain, and fatigue in women starting adjuvant letrozole treatment for breast cancer. Breast Cancer Res Treat 2010 Jan;119(1):111-118 [FREE Full text] [doi: 10.1007/s10549-009-0495-x] [Medline: 19655244]
16. Galantino ML, Desai K, Greene L, Demichele A, Stricker CT, Mao JJ. Impact of yoga on functional outcomes in breast cancer survivors with aromatase inhibitor-associated arthralgias. Integr Cancer Ther 2012 Dec;11(4):313-320 [FREE Full text] [doi: 10.1177/1534735411413270] [Medline: 21733988]
17. Kim T, Kang JW, Lee TH. Therapeutic options for aromatase inhibitor-associated arthralgia in breast cancer survivors: a systematic review of systematic reviews, evidence mapping, and network meta-analysis. Maturitas 2018 Dec;118:29-37. [doi: 10.1016/j.maturitas.2018.09.005] [Medline: 30415752]
18. Pan Y, Yang K, Shi X, Liang H, Zhang F, Lv Q. Tai chi chuan exercise for patients with breast cancer: a systematic review and meta-analysis. Evid Based Complement Alternat Med 2015;2015:535237. [doi: 10.1155/2015/535237] [Medline: 25793000]
19. Birdee GS, Wayne PM, Davis RB, Phillips RS, Yeh GY. T'ai chi and qigong for health: patterns of use in the United States. J Altern Complement Med 2009 Sep;15(9):969-973 [FREE Full text] [doi: 10.1089/acm.2009.0174] [Medline: 19757974]
20. Campo RA, O'Connor K, Light KC, Nakamura Y, Lipschitz DL, LaStayo PC, et al. Feasibility and acceptability of a Tai Chi Chih randomized controlled trial in senior female cancer survivors. Integr Cancer Ther 2013 Nov;12(6):464-474 [FREE Full text] [doi: 10.1177/1534735413485418] [Medline: 23620504]
21. Wang C, Schmid CH, Hibberd PL, Kalish R, Roubenoff R, Rones R, et al. Tai Chi is effective in treating knee osteoarthritis: a randomized controlled trial. Arthritis Rheum 2009 Nov 15;61(11):1545-1553. [doi: 10.1002/art.24832] [Medline: 19877092]
22. Monaghesh E, Hajizadeh A. The role of telehealth during COVID-19 outbreak: a systematic review based on current evidence. BMC Public Health 2020 Aug 01;20(1):1193 [FREE Full text] [doi: 10.1186/s12889-020-09301-4] [Medline: 32738884]
23. Trapani D, Marra A, Curigliano G. The experience on coronavirus disease 2019 and cancer from an oncology hub institution in Milan, Lombardy Region. Eur J Cancer 2020 Jun;132:199-206 [FREE Full text] [doi: 10.1016/j.ejca.2020.04.017] [Medline: 32380430]
24. Zhang L, Zhu F, Xie L, Wang C, Wang J, Chen R, et al. Clinical characteristics of COVID-19-infected cancer patients: a retrospective case study in three hospitals within Wuhan, China. Ann Oncol 2020 Jul;31(7):894-901 [FREE Full text] [doi: 10.1016/j.annonc.2020.03.296] [Medline: 32224151]
25. Lash BW, Katz J, Gilman P. Feasibility of a defined exercise program for aromatase inhibitor–related arthralgia (AIRA) in breast cancer survivors. J Clin Oncol 2011 May 20;29(15_suppl):e19725-e19725. [doi: 10.1200/jco.2011.29.15_suppl.e19725]
26. Western Ontario and McMaster Universities Osteoarthritis Index. Shirley Ryan AbilityLab. URL: https://www.sralab.org/rehabilitation-measures/womac-osteoarthritis-index-reliability-validity-and-responsiveness-patients [accessed 2021-11-12]
27. Swenson KK, Nissen MJ, Henly SJ, Maybon L, Pupkes J, Zwicky K, et al. Identification of tools to measure changes in musculoskeletal symptoms and physical functioning in women with breast cancer receiving aromatase inhibitors. Oncol Nurs Forum 2013 Nov;40(6):549-557. [doi: 10.1188/13.ONF.549-557] [Medline: 24161633]
28. Singer O, Cigler T, Moore AB, Levine AB, Hentel K, Belfi L, et al. Defining the aromatase inhibitor musculoskeletal syndrome: a prospective study. Arthritis Care Res (Hoboken) 2012 Dec;64(12):1910-1918. [doi: 10.1002/acr.21756] [Medline: 22730307]
29. Backhaus J, Junghanns K, Broocks A, Riemann D, Hohagen F. Test-retest reliability and validity of the Pittsburgh Sleep Quality Index in primary insomnia. J Psychosom Res 2002 Sep;53(3):737-740. [doi: 10.1016/s0022-3999(02)00330-6] [Medline: 12217446]

30. Shapiro AC, Adlis SA, Robien K, Kirstein MN, Liang S, Richter SA, et al. Randomized, blinded trial of vitamin D3 for treating aromatase inhibitor-associated musculoskeletal symptoms (AIMSS). Breast Cancer Res Treat 2016 Feb;155(3):501-512 [FREE Full text] [doi: 10.1007/s10549-016-3710-6] [Medline: 26868123]
31. Hershman DL, Unger JM, Crew KD, Awad D, Dakhil SR, Gralow J, et al. Randomized multicenter placebo-controlled trial of omega-3 fatty acids for the control of aromatase inhibitor-induced musculoskeletal pain: SWOG S0927. J Clin Oncol 2015 Jun 10;33(17):1910-1917 [FREE Full text] [doi: 10.1200/JCO.2014.59.5595] [Medline: 25940724]
32. Angst F, Aeschlimann A, Stucki G. Smallest detectable and minimal clinically important differences of rehabilitation intervention with their implications for required sample sizes using WOMAC and SF-36 quality of life measurement instruments in patients with osteoarthritis of the lower extremities. Arthritis Rheum 2001 Aug;45(4):384-391. [doi: 10.1002/1529-0131(200108)45:4<384::AID-ART352>3.0.CO;2-0] [Medline: 11501727]
33. Chen S, Feng Z, Yi X. A general introduction to adjustment for multiple comparisons. J Thorac Dis 2017 Jun;9(6):1725-1729. [doi: 10.21037/jtd.2017.05.34] [Medline: 28740688]
34. Trevino KM, Raghunathan N, Latte-Naor S, Polubriaginof FCG, Jensen C, Atkinson TM, et al. Rapid deployment of virtual mind-body interventions during the COVID-19 outbreak: feasibility, acceptability, and implications for future care. Support Care Cancer 2021 Feb;29(2):543-546 [FREE Full text] [doi: 10.1007/s00520-020-05740-2] [Medline: 32902712]

Abbreviations

AI: aromatase inhibitor
AUSCAN: Australian Canadian Osteoarthritis Hand Index
BC: breast cancer
BPI: Brief Pain Inventory
CES-D: Center for Epidemiological Studies–Depression
FSI: Fatigue Symptom Inventory
HFRDIS: Hot Flash Related Daily Interference Scale
PSQI: Pittsburgh Sleep Quality Index
RCT: randomized clinical trial
WOMAC: Western Ontario and McMaster University Osteoarthritis Index

Edited by A Mavragani; submitted 15.11.21; peer-reviewed by K Crew; comments to author 10.12.21; revised version received 22.02.22; accepted 16.05.22; published 21.06.22.

Please cite as:
Gomaa S, West C, Lopez AM, Zhan T, Schnoll M, Abu-Khalaf M, Newberg A, Wen KY
A Telehealth-Delivered Tai Chi Intervention (TaiChi4Joint) for Managing Aromatase Inhibitor–Induced Arthralgia in Patients With Breast Cancer During COVID-19: Longitudinal Pilot Study
JMIR Form Res 2022;6(6):e34995
URL: https://formative.jmir.org/2022/6/e34995
doi:10.2196/34995
PMID:35727609

Original Paper

The AI Will See You Now: Feasibility and Acceptability of a Conversational AI Medical Interviewing System

Grace Hong[1], BA; Margaret Smith[1], MBA; Steven Lin[1], MD

Stanford Healthcare AI Applied Research Team, Division of Primary Care and Population Health, Stanford University School of Medicine, Redwood City, CA, United States

Corresponding Author:
Grace Hong, BA
Stanford Healthcare AI Applied Research Team
Division of Primary Care and Population Health
Stanford University School of Medicine
585 Broadway
Suite 800
Redwood City, CA, 94063
United States
Phone: 1 847 800 1377
Email: hongrace@stanford.edu

Abstract

Background: Primary care physicians (PCPs) are often limited in their ability to collect detailed medical histories from patients, which can lead to errors or delays in diagnosis. Recent advances in artificial intelligence (AI) show promise in augmenting current human-driven methods of collecting personal and family histories; however, such tools are largely unproven.

Objective: The main aim of this pilot study was to evaluate the feasibility and acceptability of a conversational AI medical interviewing system among patients.

Methods: The study was conducted among adult patients empaneled at a family medicine clinic within a large academic medical center in Northern California. Participants were asked to test an AI medical interviewing system, which uses a conversational avatar and chatbot to capture medical histories and identify patients with risk factors. After completing an interview with the AI system, participants completed a web-based survey inquiring about the performance of the system, the ease of using the system, and attitudes toward the system. Responses on a 7-point Likert scale were collected and evaluated using descriptive statistics.

Results: A total of 20 patients with a mean age of 50 years completed an interview with the AI system, including 12 females (60%) and 8 males (40%); 11 were White (55%), 8 were Asian (40%), and 1 was Black (5%), and 19 had at least a bachelor's degree (95%). Most participants agreed that using the system to collect histories could help their PCPs have a better understanding of their health (16/20, 80%) and help them stay healthy through identification of their health risks (14/20, 70%). Those who reported that the system was clear and understandable, and that they were able to learn it quickly, tended to be younger; those who reported that the tool could motivate them to share more comprehensive histories with their PCPs tended to be older.

Conclusions: In this feasibility and acceptability pilot of a conversational AI medical interviewing system, the majority of patients believed that it could help clinicians better understand their health and identify health risks; however, patients were split on the effort required to use the system, and whether AI should be used for medical interviewing. Our findings suggest areas for further research, such as understanding the user interface factors that influence ease of use and adoption, and the reasons behind patients' attitudes toward AI-assisted history-taking.

(JMIR Form Res 2022;6(6):e37028) doi:10.2196/37028

KEYWORDS

artificial intelligence; feasibility studies; patient acceptance of health care; diagnostic errors; patient-generated health data; clinical; medical history; healthcare; health care

Introduction

Primary care providers (PCPs) face numerous challenges, including time constraints and burnout, and are often limited in their ability to collect detailed medical histories from their patients [1]. Currently, patients' personal and family medical histories are collected via paper forms or interviewing prior to and/or during a patient visit, with manual data entry into the electronic health record (EHR). Often, data are missing or of poor quality due to lack of time or lack of training [2].

Information gaps can lead to errors or delays in diagnosis and failures to address actionable risk factors, which affect an estimated 12 million Americans each year [3]. Diagnosis in primary care is a high-risk area for errors, for several reasons. PCPs typically face high patient volumes, make decisions amid uncertainty, and must balance the risk of missed or delayed diagnoses with the stewardship of scarce resources [4]. Poor communication leading to gaps in information sharing is a major driver of diagnostic errors in primary care settings [4], while engaging and empowering patients in the task of generating data using technology aims to close those gaps and support personalized medicine [5].

Recent advances in conversational agents powered by artificial intelligence (AI) and natural language processing show promise in augmenting current human-driven methods of collecting personal and family histories as part of pre-visit planning [6]. Various types of conversational agents have emerged, including chatbots, embodied conversational agents (avatars), and voice assistants, all of which mimic human conversation using text and/or spoken language [7]. Within health care, these agents have been used for facilitating screening for health conditions, triage, counseling, self-management of chronic conditions, and training for health care professionals; reported benefits have included their potential to support populations with poor health literacy, be scaled to reach large populations, and improve patient engagement [8]. However, such tools remain largely unproven in real-world settings, and few studies have assessed the use case of collecting patients' information before an appointment to provide tailored counseling [2,9-11]. This study aimed to assess the feasibility and acceptability of a conversational AI medical interviewing system from the perspective of primary care patients.

Methods

Setting and Participants

The study was conducted between February and April 2021 at a family medicine clinic within a large academic medical center in Northern California. Participants were eligible to participate if they were aged ≥18 years and English-speaking. Research staff contacted eligible participants via email and telephone, providing a brief summary of the study and asking whether they would be interested in participating.

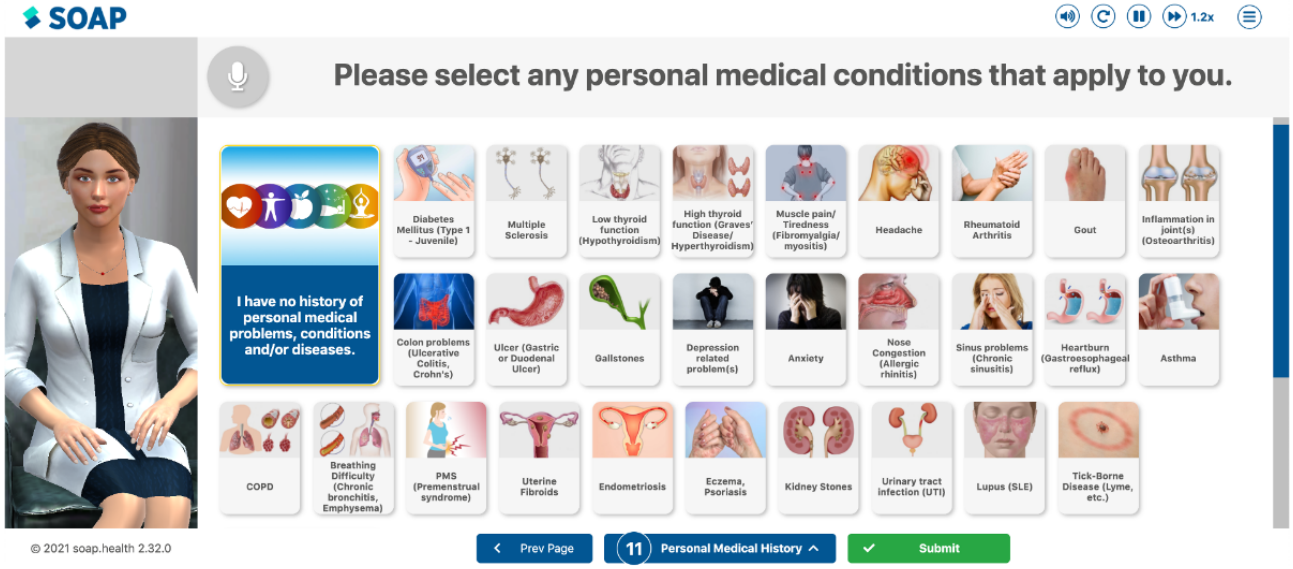
Ethical Considerations

The Stanford University Institutional Review Board reviewed this study and exempted it (protocol number IRB-59555).

Procedure

Participants were asked to test the web-based AI medical interviewing system on a personal computer at a time and location of their choosing. The program developed by SOAP Health uses AI and natural language processing to convert speech to text and provide appropriate responses to user-entered data through Genie, a user-facing conversational avatar and chatbot. Genie asked participants a series of questions to (1) capture detailed personal medical histories, multigenerational family histories, and social determinants of health data (eg, financial insecurity, food insecurity, access to affordable health care, access to transportation); and (2) identify risk factors based on established guidelines for further evaluation (eg, hereditary cancers, cardiovascular disease), based on personal or family histories (Figure 1 and Multimedia Appendix 1). Questions covered 25 topic areas and were both asked aloud by Genie and displayed visually onscreen (Multimedia Appendix 1). Participants were able to respond by speaking or clicking a displayed response option and were allowed to use as much time as needed for the interview; most completed it in 30-45 minutes.

Figure 1. Screenshot of Genie and the conversational AI medical interviewing system. AI: artificial intelligence.



Survey and Outcomes

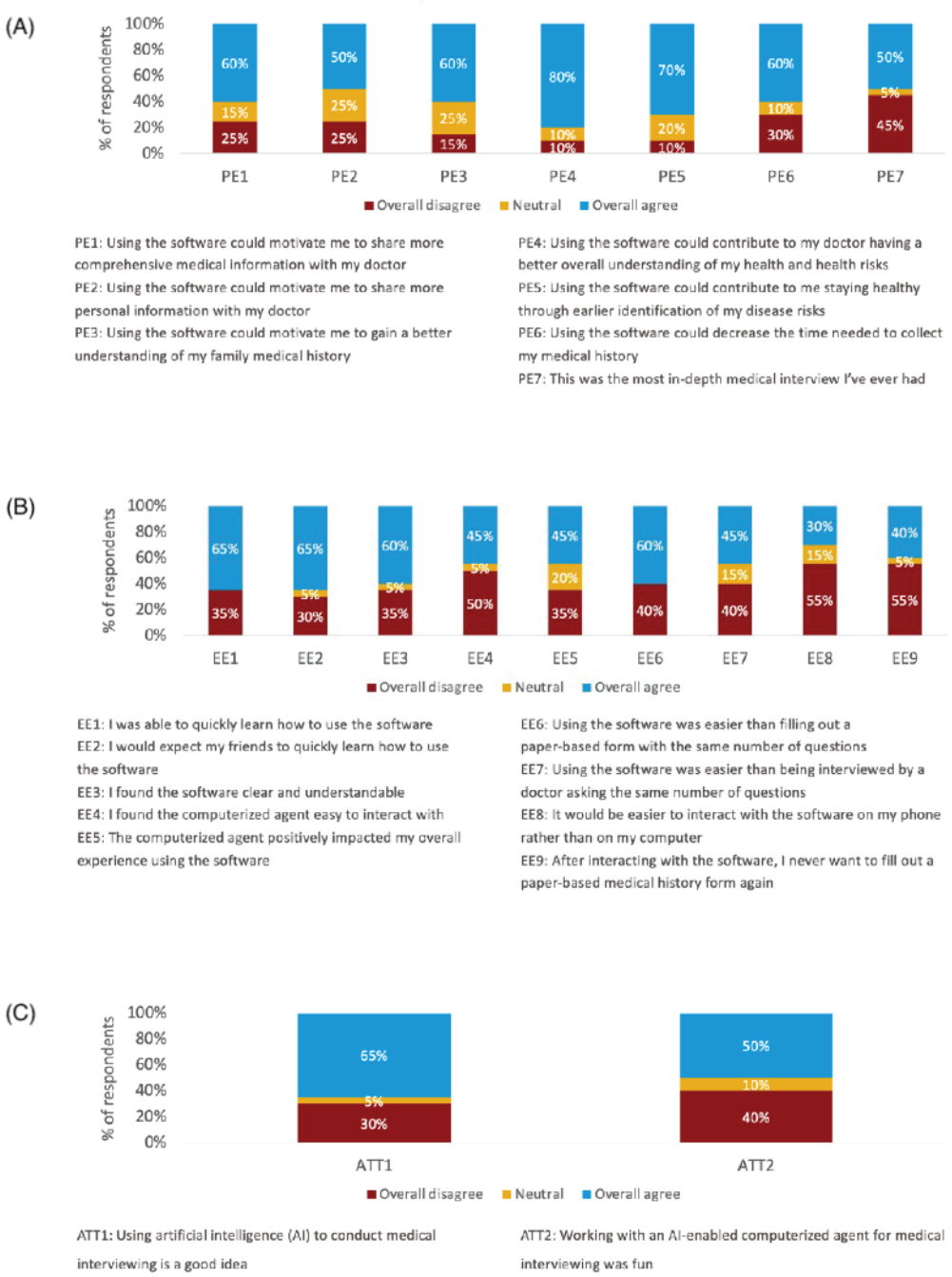
After completing an interview with the AI system, participants were asked to complete a web-based survey based on the validated Unified Theory of Acceptance and Use of Technology framework [12], a model explaining user acceptance and adoption of new technology that has been used across various studies assessing emerging technology [13-15]. The outcomes were patient-reported feasibility and acceptability ratings from the survey based on responses to Likert-scale questions in the following domains: (1) performance expectancy, the degree to which patients believe using the system will help them share relevant information with their PCPs and identify disease risks earlier; (2) effort expectancy, the degree of ease patients associate with using the system; and (3) attitude toward using technology, the degree to which patients have a positive attitude toward using the system. Participants provided responses on a 7-point scale (strongly disagree to strongly agree) to statements including "Using the software could contribute to my doctor having a better overall understanding of my health and health risks," "I was able to quickly learn how to use the software," and "Using artificial intelligence to conduct medical interviewing is a good idea." The survey also collected patients' demographic information including their age, gender, race/ethnicity, and highest level of education. Descriptive statistics were used to compute counts with percentages for survey responses.

Results

Twenty patients with a mean age of 50 years completed an interview with the AI system, including 12 females (60%) and 8 males (40%); 11 were White (55%), 8 were Asian (40%), 1 was Black (5%), and 19 had at least a bachelor's degree (95%) (Multimedia Appendix 1).

The majority of participants agreed that using a conversational AI medical interviewing system to collect histories could help PCPs have a better understanding of their health (16/20, 80%) and help them stay healthy through identification of their health risks (14/20, 70%) (Figure 2). Participants who felt the system was easy to use tended to be younger. The median age for those who agreed that they were able to learn the system quickly was 41 years compared to 61 years for those who disagreed and 41 years for those who agreed that the system was clear and understandable compared to 68 years for those who disagreed (Multimedia Appendix 1). Those who reported that the tool could motivate them to share more comprehensive medical information with their PCPs tended to be older—median age of 52 years for those who agreed and 36 years for those who disagreed (Multimedia Appendix 1). Patients were split on the effort required to use the tool, and on whether AI should be used for medical interviewing (Figure 2).

Figure 2. Patient-reported feasibility and acceptability ratings of the system. AI: artificial intelligence; PE: performance expectancy; EE: effort expectancy; ATT: attitude toward using technology.



Discussion

In this feasibility and acceptability pilot of a conversational AI medical interviewing system, the majority of primary care patients believed that such a tool could help PCPs better understand their health and identify their health risks. However, while performance expectancy of the system was predominantly positive, results were mixed in terms of effort expectancy and attitude toward the emerging technology.

Our results are aligned with existing literature. A recent systematic review examining the effectiveness and usability of AI-based conversational agents in health care found that 67% of the 31 studies reported positive or mixed evidence supporting the effectiveness, usability, and positive user perceptions of the agents [8]. Additionally, in a study exploring the acceptability and feasibility of a virtual counselor to collect family health histories in an underserved population, a vast majority of participants found the virtual counselor easy to use and understood the questions being asked [2]. At the same time, studies reporting qualitative feedback have consistently cited the following as barriers that will need to be addressed before conversational agents can be deployed and used widely: agents having difficulty understanding users, agents being repetitive and not sufficiently interactive, and users having difficulty forming connections with the agent [8]. Our study adds to this

growing body of evidence by assessing the use of conversational agents in a slightly older population in primary care.

Our study has several limitations including the small sample size and convenience sampling recruitment approach. Those who opted to participate may have had more positive notions about new technologies and may have evaluated the system more favorably than the general population. Moreover, participants in our study were highly educated, which may limit the generalizability of our findings.

The use of conversational agents to gather more detailed information about patients' personal medical history, family medical history, and social determinants of health could aid in capturing a more holistic view of patients and identifying disease risks earlier. Our findings suggest areas for further research, such as understanding the user interface factors that influence ease of use and adoption, and the reasons behind patients' dichotomous attitudes toward AI-assisted history-taking.

Acknowledgments

The authors thank Bridget Scallen for her research support.

This project was jointly funded by SOAP Health (Boca Raton, FL) and the Division of Primary Care and Population Health at Stanford University School of Medicine. SOAP Health had no role in the preparation of this manuscript and the decision to approve publication of the finished product.

Prior Presentations

This project was presented at the North American Primary Care Research Group (NAPCRG) Annual Virtual Meeting (November 19-23, 2021) and the Society of Teachers of Family Medicine (STFM) Annual Spring Conference (April 30-May 4, 2022).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional information about the artificial intelligence medical interviewing system; demographics of participants; and additional information on patient-reported feasibility and acceptability ratings of the system.
[DOCX File , 1082 KB - formative_v6i6e37028_app1.docx]

References

1. National Academies of Sciences, Engineering, and Medicine. In: Balogh EP, Miller BT, Ball JR, editors. Improving diagnosis in health care. Washington, DC: The National Academies Press; 2015.
2. Wang C, Bickmore T, Bowen DJ, Norkunas T, Campion M, Cabral H, et al. Acceptability and feasibility of a virtual counselor (VICKY) to collect family health histories. Genet Med 2015;17(10):822-830 [FREE Full text] [doi: 10.1038/gim.2014.198] [Medline: 25590980]
3. Singh H, Meyer A, Thomas EJ. The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations. BMJ Qual Saf 2014;23(9):727-731 [FREE Full text] [doi: 10.1136/bmjqs-2013-002627] [Medline: 24742777]
4. Singh H, Schiff GD, Graber ML, Onakpoya I, Thompson MJ. The global burden of diagnostic errors in primary care. BMJ Qual Saf 2017 Jun 16;26(6):484-494 [FREE Full text] [doi: 10.1136/bmjqs-2016-005401] [Medline: 27530239]
5. Tiase V, Hull W, McFarland M, Sward KA, Del Fiol G, Staes C, et al. Patient-generated health data and electronic health record integration: a scoping review. JAMIA Open 2020 Dec;3(4):619-627 [FREE Full text] [doi: 10.1093/jamiaopen/ooaa052] [Medline: 33758798]
6. Holdsworth LM, Park C, Asch SM, Lin S. Technology-enabled and artificial intelligence support for pre-visit planning in ambulatory care: findings from an environmental scan. Ann Fam Med 2021 Sep 13;19(5):419-426 [FREE Full text] [doi: 10.1370/afm.2716] [Medline: 34546948]
7. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 01;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
8. Milne-Ives M, de Cock C, Lim E, Shehadeh MH, de Pennington N, Mole G, et al. The effectiveness of artificial intelligence conversational agents in health care: systematic review. J Med Internet Res 2020 Oct 22;22(10):e20346 [FREE Full text] [doi: 10.2196/20346] [Medline: 33090118]
9. Adams WG, Phillips BD, Bacic JD, Walsh KE, Shanahan CW, Paasche-Orlow MK. Automated conversation system before pediatric primary care visits: a randomized trial. Pediatrics 2014 Sep 04;134(3):e691-e699. [doi: 10.1542/peds.2013-3759] [Medline: 25092938]

10. Ponathil A, Ozkan F, Bertrand J, Agnisarman S, Narasimha S, Welch B, et al. An empirical study investigating the user acceptance of a virtual conversational agent interface for family health history collection among the geriatric population. Health Informatics J 2020 Dec 16;26(4):2946-2966 [FREE Full text] [doi: 10.1177/1460458220955104] [Medline: 32938275]
11. Gashi F, Regli S, May R, Tschopp P, Denecke K. Developing intelligent interviewers to collect the medical history: lessons learned and guidelines. In: Navigating healthcare through challenging times. Amsterdam: IOS Press Ebooks; 2021:18-25.
12. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Quarterly 2003;27(3):425-478. [doi: 10.2307/30036540]
13. Hoogenbosch B, Postma J, de Man-van Ginkel JM, Tiemessen NA, van Delden JJ, van Os-Medendorp H. Use and the users of a patient portal: cross-sectional study. J Med Internet Res 2018 Sep 17;20(9):e262 [FREE Full text] [doi: 10.2196/jmir.9418] [Medline: 30224334]
14. Janssen A, Keep M, Selvadurai H, Kench A, Hunt S, Simonds S, et al. Factors that influence use of a patient portal by health professionals. Int J Environ Res Public Health 2021 Feb 15;18(4):1877 [FREE Full text] [doi: 10.3390/ijerph18041877] [Medline: 33671953]
15. Apolinário-Hagen J, Menzel M, Hennemann S, Salewski C. Acceptance of mobile health apps for disease management among people with multiple sclerosis: web-based survey study. JMIR Form Res 2018 Dec 12;2(2):e11977 [FREE Full text] [doi: 10.2196/11977] [Medline: 30684408]

Abbreviations

AI: artificial intelligence
PCP: primary care provider

Edited by A Mavragani; submitted 03.02.22; peer-reviewed by W Liaw, M Elbattah, M Sillice; comments to author 16.03.22; revised version received 09.06.22; accepted 13.06.22; published 27.06.22.

Please cite as:
Hong G, Smith M, Lin S
The AI Will See You Now: Feasibility and Acceptability of a Conversational AI Medical Interviewing System
JMIR Form Res 2022;6(6):e37028
URL: https://formative.jmir.org/2022/6/e37028
doi:10.2196/37028
PMID:35759326

Original Paper

A Mobile Education and Social Support Group Intervention for Improving Postpartum Health in Northern India: Development and Usability Study

Alison M El Ayadi[1,2], MPH, SCD; Mona Duggal[3], MD; Rashmi Bagga[4], MD; Pushpendra Singh[5], PhD; Vijay Kumar[6], MD; Alka Ahuja[4], MD; Ankita Kankaria[7], MD; Darshan Hosapatna Basavarajappa[3], MD; Jasmeet Kaur[5], MTech; Preetika Sharma[3], PhD; Swati Gupta[5], MS; Ruchita S Pendse[8], BA; Laura Weil[1], MSN, MPH; Dallas Swendeman[9], MPH, PhD; Nadia G Diamond-Smith[2], MS, PhD

[1]Department of Obstetrics, Gynecology and Reproductive Sciences, University of California, San Francisco, San Francisco, CA, United States

[2]Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA, United States

[3]Postgraduate Institute of Medical Education & Research, Chandigarh, India

[4]Department of Obstetrics & Gynaecology, Postgraduate Institute of Medical Education & Research, Chandigarh, India

[5]Department of Computer Science & Engineering, Indraprastha Institute of Information Technology Delhi, New Delhi, India

[6]Survival for Women and Children Foundation, Panchkula, India

[7]Department of Community and Family Medicine, All India Institute of Medical Sciences, Bathinda, India

[8]School of Medicine, Stanford University, Palo Alto, CA, United States

[9]Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles, Los Angeles, CA, United States

Corresponding Author:
Alison M El Ayadi, MPH, SCD
Department of Obstetrics, Gynecology and Reproductive Sciences
University of California, San Francisco
550 16th Street
San Francisco, CA, 94158
United States
Phone: 1 4156598367
Email: alison.elayadi@ucsf.edu

Abstract

Background: Structural and cultural barriers limit Indian women's access to adequate postnatal care and support despite their importance for maternal and neonatal health. Targeted postnatal education and support through a mobile health intervention may improve postnatal recovery, neonatal care practices, nutritional status, knowledge and care seeking, and mental health.

Objective: We sought to understand the feasibility and acceptability of our first pilot phase, a flexible 6-week postnatal mobile health intervention delivered to 3 groups of women in Punjab, India, and adapt our intervention for our next pilot phase, which will formally assess intervention feasibility, acceptability, and preliminary efficacy.

Methods: Our intervention prototype was designed to deliver culturally tailored educational programming via a provider-moderated, voice- and text-based group approach to connect new mothers with a social support group of other new mothers, increase their health-related communication with providers, and refer them to care needed. We targeted deployment using feature phones to include participants from diverse socioeconomic groups. We held moderated group calls weekly, disseminated educational audios, and created SMS text messaging groups. We varied content delivery, group discussion participation, and chat moderation. Three groups of postpartum women from Punjab were recruited for the pilot through community health workers. Sociodemographic data were collected at baseline. Intervention feasibility and acceptability were assessed through weekly participant check-ins (N=29), weekly moderator reports, structured end-line in-depth interviews among a subgroup of participants (15/29, 52%), and back-end technology data.

Results: The participants were aged 24 to 28 years and 1 to 3 months postpartum. Of the 29 participants, 17 (59%) had their own phones. Half of the participants (14/29, 48%) attended ≥3 of the 6 calls; the main barriers were childcare and household responsibilities and network or phone issues. Most participants were very satisfied with the intervention (16/19, 84%) and found the educational content (20/20, 100%) and group discussions (17/20, 85%) very useful. The participants used the SMS text

messaging chat, particularly when facilitator-moderated. Sustaining participation and fostering group interactions was limited by technological and sociocultural challenges.

Conclusions: The intervention was considered generally feasible and acceptable, and protocol adjustments were identified to improve intervention delivery and engagement. To address technological issues, we engaged a cloud-based service provider for group calls and an interactive voice response service provider for educational recordings and developed a smartphone app for the participants. We seek to overcome sociocultural challenges through new strategies for increasing group engagement, including targeting midlevel female community health care providers as moderators. Our second pilot will assess intervention feasibility, acceptability, and preliminary effectiveness at 6 months. Ultimately, we seek to support the health and well-being of postpartum women and their infants in South Asia and beyond through the development of efficient, acceptable, and effective intervention strategies.

(JMIR Form Res 2022;6(6):e34087) doi:10.2196/34087

KEYWORDS

mHealth; group care; postpartum; postnatal; antenatal; India; pilot; mobile phone

Introduction

Background

Against the backdrop of persistent rural poverty in Northern India, modest gains in maternal health have been made over the past decade, with maternal mortality decreasing from 251 to 113 maternal deaths per 100,000 live births [1,2]. Despite this progress, further improvements in maternal health are needed for India to achieve the United Nations Sustainable Development Goal target of <70 maternal deaths per 100,000 live births by 2030 [3], including a broader focus on achieving the full continuum of perinatal care. The COVID-19 pandemic has significantly impeded global progress in achieving Sustainable Development Goal targets by exacerbating maternal and perinatal adversity, with significantly greater impacts seen among low- and middle-income countries such as India compared with higher-income settings [4].

Postnatal care and support are important contributors to optimizing maternal and neonatal health and well-being and a critical component of the full continuum of perinatal care [5]. High-quality postnatal care and social support (ie, emotional, instrumental, and informational) have been associated with reduced maternal and neonatal mortality [5,6] and increased maternal engagement in behaviors promoting newborn (eg, exclusive breastfeeding and child immunization) and maternal (eg, postnatal adoption of family planning) health [7,8]. Acknowledging the important role of postnatal care, the Government of India recommends that women receive 3 postnatal visits from community health workers [1]. However, despite India's broad community-based maternal health programming, a significant drop in the continuum of peripartum care occurs in the postnatal period [9,10]. Postnatal care achievement is generally low across the country, with 65% of Indian women receiving at least one health check within 2 days of birth, ranging from 48% to 80% across wealth quintiles, and full postnatal care achievement being anecdotally low [1].

Substantial structural and cultural barriers prevent new mothers from attending postnatal care at facilities or other locations that may be far from their homes, particularly in India. Common logistical challenges are exacerbated in India by rural geographic distances, cultural and linguistic barriers to care, women's practice of postnatal seclusion, and generally low levels of mobility in marriage, all culminating in reduced care access [11-13]. Further intergenerational and gender-based hierarchical roles structure decision-making in Indian households, particularly for couples living in extended-family households where decision-making may be largely outside the hands of women [14]. Within this context, novel methods for improving women's access to the full peripartum continuum of health care are required.

Mobile phone–based health approaches (mobile health [mHealth]) offer innovative opportunities to overcome logistical barriers to postnatal care access. Despite women's physical mobility limitations in this setting, 88% of Indian households nationally own a mobile phone, including 72% of women [1,15]. Theory-informed mobile support and education groups have positively affected [16] maternal and neonatal health outcomes [17-20], and mHealth approaches to pregnancy care suggest high acceptability, promising results, and high cost-effectiveness [21,22]. However, preventive mHealth models extending care postnatally are sparse [23-25]. Group-participatory learning and action models (eg, women's groups) are another promising model for improving postnatal care given their efficient approach, acceptability, and effectiveness in improving maternal and neonatal health indicators [26,27]. Such models provide efficient health education, strengthen social support networks, and influence social norms [28]. Social support [29,30] reduces postnatal depression and anxiety [31,32] and improves breastfeeding maintenance [33,34] and women's empowerment [35]. CenteringPregnancy and other group-based supportive prenatal care models affect exclusive breastfeeding, contraceptive uptake, depression, and immunization postnatally [36-40]. Extending group care postnatally is an innovative and promising approach for improving health outcomes for both mothers and neonates through targeted education and support for women in the postnatal period and referral to in-person care.

Objectives

This study describes the development and pilot results of a targeted education and support intervention for postnatal Indian women where they can interact with providers and other new mothers, named *Maa Shishu Swasthya Sahayak Samooh* (maternal and child health support group; *MeSSSSage*). With the goal of optimizing the feasibility and acceptability of our

intervention, we implemented a 2-phase developmental process to pilot test and refine the intervention functions, processes, and delivery platform. In this paper, we report on the results of pilot phase 1, a flexible 6-week postnatal mHealth intervention delivered to 3 groups of women in Punjab, India, and describe the integration of these results into optimizing the intervention for our phase 2 pilot study, in which we will seek to assess the feasibility and acceptability of the full 6-month intervention more formally.

Methods

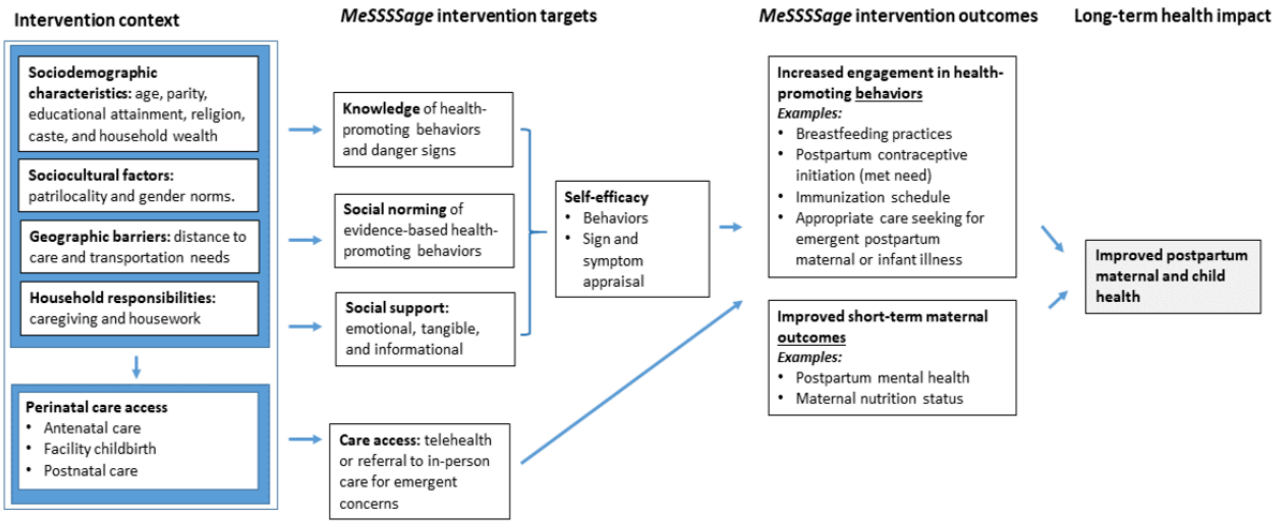
Study Context

Maternal and child health indicators in Punjab, Northern India, show some improvement, but persistent challenges remain, suggesting that innovative strategies are needed for improving health outcomes and ensuring high-quality care. Across the perinatal continuum of care, 68.5% of new mothers receive at least four antenatal care visits and 90.5% have facility births, but only 40% of women receive any postnatal care [1]. Many Punjabi women of reproductive age are anemic (53.5%), and anemia rates are higher among postpartum women [41]. Modern contraceptive use is 63% overall [42], yet postpartum uptake of contraceptives is lower. A study found that only 30% of Punjab women adopted a contraceptive method postpartum, most often condoms, with 16% having an unmet need [43]. Postpartum mental health remains understudied; however, a recent systematic review of postpartum depression in India reported a pooled estimate of 22% and identified lack of social support as a primary risk factor [44]. Most children aged 12 to 23 months (89.1%) are fully immunized; however, appropriate diarrheal disease and acute respiratory infection management in children aged <5 years is inadequate (range 26.7%-90.3%) [42]. Only 5% of children aged <2 years receive adequate nutrition, and approximately one-fifth of children aged <5 years are stunted (25.7%) or underweight (21.6%) [42]. Disparities in maternal and child health indicators exist according to rurality of residence and socioeconomic status [42].

Intervention Prototype Development

The initial *MeSSSSage* prototype was developed by a team of Indian and US-based maternal health clinicians, researchers, and human-centered technology design experts. The prototype was based on the interdisciplinary literature, early-stage formative needs assessment research with postnatal women, and local and international care guidelines. *MeSSSSage* was designed to overcome the prevalent structural and cultural barriers to postnatal maternal and neonatal health care in this setting by delivering culturally tailored educational programming via a provider-moderated, voice- and text-based group approach to connect mothers with a social support group of other new mothers, increase their health-related communication with providers, and refer them in a timely and appropriate manner to care. Our conceptual framework (Figure 1) outlines a summary model of the factors identified in our formative work that influence perinatal care access, knowledge, health behaviors, and health outcomes, and the pathways that the *MeSSSSage* intervention has been designed to disrupt. The intervention is registered at ClinicalTrials.gov (NCT04636398).

Figure 1. Conceptual framework of Maa Shishu Swasthya Sahayak Samooh (maternal and child health support group; *MeSSSSage*) intervention context, targets, outcomes, and anticipated health impacts.

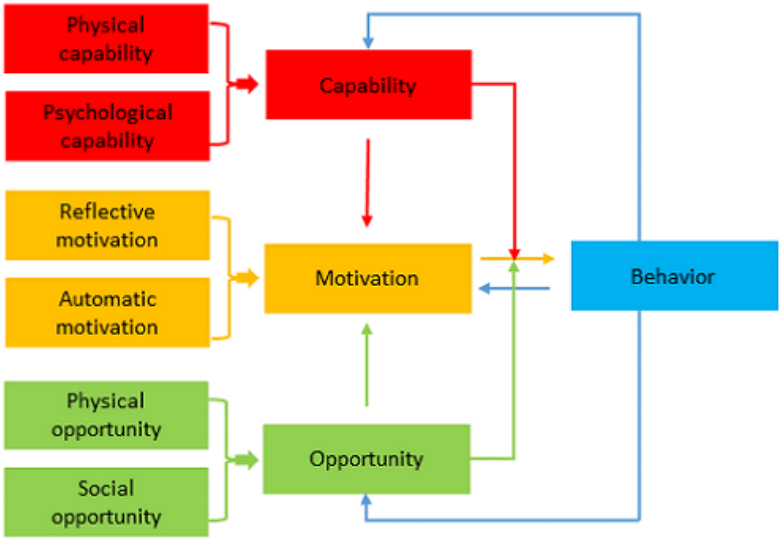


Intervention Conceptual Framework

Capability, Opportunity, Motivation-Behavior is a comprehensive framework for behavior change intervention design that posits that behaviors (B) are the result of interactions between capability (C), opportunity (O), and motivation (M; Figure 2) [45,46]. The *MeSSSSage* intervention builds on the Capability, Opportunity, Motivation-Behavior framework to affect health behaviors as follows: (1) the mHealth approach allows women to participate in the intervention from their homes to reduce the impact of social, sociocultural, geographic, and logistical barriers, increasing *physical and social opportunity*; (2) the educational content is designed to increase knowledge of important maternal and child health–promoting behaviors, increasing *psychological capability* and *reflective motivation*; and (3) the social support and social norming component is designed to increase *reflective and automatic motivation.* We hypothesize that providing targeted education and support to

women in the postnatal period through a mobile social network where they can interact with providers and other new mothers will improve their knowledge of health-promoting behaviors and their parental self-efficacy and empowerment. By appropriately identifying infant and maternal danger signs and promoting timely care seeking for routine and emergency visits (including referrals), we seek to reduce both maternal and infant mortality and morbidity. Furthermore, mHealth group support could help encourage and sustain exclusive breastfeeding and improve uptake of postnatal family planning and childhood vaccination, which positively affects maternal and child health in both the short and long term.

Figure 2. Capability, Opportunity, Motivation-Behavior model of behavior change [45,46].



Intervention Educational Content

Maternal and newborn educational content was developed through a review of Indian and international clinical guidelines, consultation with experienced local maternal and child health providers, and formative research to identify health education needs among our target population. Topics were organized according to anticipated maternal and neonatal information needs by postpartum week (Table 1). The discussion facilitated during the group calls focused on ensuring women's understanding of the content and addressing questions. An obstetrician or gynecologist and a pediatrician advised on emergent health concerns and referred women for in-person care.

Table 1. Maternal and newborn educational content, Maa Shishu Swasthya Sahayak Samooh (maternal and child health support group) pilot intervention, phase 1.

Intervention week	Maternal educational topics	Newborn educational topics
1	Breastfeeding, postpartum hygiene, and self-care	Infant danger signs
2	Diet and nutrition and maternal danger signs	Temperature regulation
3	Mental health and COVID-19 precautions	Neonatal care
4	Emotional support and family planning overview	Massage
5	Family planning methods and postnatal care visits	Developmental milestones
6	General health issues and use of mobile health technology for health improvement	COVID-19 precautions

Technology Development

For this phase 1 pilot, our team prioritized deployment across feature (nonsmart) phones for inclusivity of diverse groups of women, including those of lower socioeconomic status. We developed 2 platforms for group call implementation and the dissemination of educational audios using FreeSWITCH open-source, voice-over IP servers (Multimedia Appendix 1). Group calls initiated from the server were free for intervention participants; however, accessing educational audio recordings outside of the weekly calls incurred call charges to the participants. All intervention content was in the Punjabi language.

A mobile SMS text messaging group that included participants and moderators was created for each intervention group using WhatsApp (Meta Platforms). This SMS text messaging group was used to share group call reminders and educational material and to asynchronously provide additional education and support to the intervention participants. The participants were invited to use the WhatsApp group to share stories or experiences with their group members and to ask questions of other participants and moderators. Some discussions were facilitated by the group moderator per our intervention testing plan (Table 2). The WhatsApp groups were accessible to all participants with smartphones and many participants with feature phones.

Table 2. Intervention delivery approaches by domain, Maa Shishu Swasthya Sahayak Samooh (maternal and child health support group) pilot intervention, phase 1.

Week and domain	Group 1	Group 2	Group 3
Week 1			
Educational component delivery[a]	Live	Live	Recorded
Participation[b]	Talk	Hand	Talk
SMS text messaging group moderation[c]	Not moderated	Moderated	Moderated
Week 2			
Educational component delivery[a]	Live	Live	Recorded
Participation[b]	Talk	Hand	Talk
SMS text messaging group moderation[c]	Not moderated	Moderated	Moderated
Week 3			
Educational component delivery[a]	Live	Live	Recorded
Participation[b]	Hand	Talk	Hand
SMS text messaging group moderation[c]	Not moderated	Moderated	Moderated
Week 4			
Educational component delivery[a]	Recorded	Recorded	Live
Participation[b]	Hand	Talk	Hand
SMS text messaging group moderation[c]	Moderated	Not moderated	Not moderated
Week 5			
Educational component delivery[a]	Recorded	Recorded	Live
Participation[b]	Talk	Hand	Talk
SMS text messaging group moderation[c]	Moderated	Not moderated	Not moderated
Week 6			
Educational component delivery[a]	Recorded	Recorded	Live
Participation[b]	Talk	Hand	Talk
SMS text messaging group moderation[c]	Moderated	Not moderated	Not moderated

[a]For *live*, educational messages are delivered by the moderator during the group call, intermixing the material with open time for questions. For *recorded*, prerecorded educational messages are accessed via phone before the group call, with the full group call used for discussion.

[b]For *hand*, the women have to raise their hands to talk, with the moderator unmuting them to allow for their participation. For *talk*, the women can talk freely.

[c]For *moderated*, a group moderator facilitates group participation through consistent use of prompts in the group. For *not moderated*, the SMS text messaging group is introduced to the women for their own use with no researcher facilitation.

Intervention Delivery

The pilot intervention was delivered over a 6-week period (December 2020 to January 2021). This was between the first and second COVID-19 surges in India and, thus, although some restrictions were in place, COVID-19 rates were low, and care access was not seriously limited. Weekly group calls were planned for approximately 30 to 40 minutes and scheduled at the same time each week. SMS text message reminders were sent to the participants 1 day beforehand. Participation in the WhatsApp chat group was encouraged. Group calls were moderated by 3 research team members: a moderator to handle group dynamics (AA), an obstetrician or gynecologist (RB or DHB), or a pediatrician (VK).

Varying Delivery Approaches

To determine the best structure and understand how to optimize engagement, we varied three delivery approaches to assess the women's reception of delivery of educational content, participation in group discussion, and format of the SMS text messaging–based chat platform (Table 2): (1) the educational content being provided live by the moderator versus through a recorded message delivered through interactive voice response before the group call, (2) participation in calls occurring via the

participants *raising their hand* by pressing 1 on their keypad to indicate that they would like to speak and being unmuted by the technical moderator versus all participants being unmuted for the full call, and (3) the WhatsApp SMS text messaging group being moderated by a provider through consistent use of prompts in the group versus not being moderated (both groups could text questions in the chat for the provider to answer). Each group experienced each of these modalities for half of the intervention.

Study Site and Participants

The study was conducted within Block Boothgarh, Mohali District, Punjab state, Northern India. Block Boothgarh comprises 129 villages with a total population of 5000 served by 16 health subcenters, each with 1 auxiliary nurse midwife (ANM) who maintains a consolidated perinatal care register. Mobile phone ownership per local care registers is high among households (90%) and women (50%).

Study participants included women residing in the study area who had given birth within the previous 3 months. Further inclusion criteria were being aged <40 years and having a live neonate with birth weight >1500 g. The exclusion criteria were maternal complications during or after childbirth warranting hospital stay and continued medical care at the facility, stillbirth, twins, significant birth defects, inability to provide informed consent, and lack of phone access if unwilling to accept a phone from the study team. If the eligible participants did not have a personal phone, they were offered a study phone.

Potential participants were identified with assistance from local ANMs. Women who met the study inclusion criteria based on ANM records were contacted telephonically by a study researcher who further screened them and explained the study procedures in detail, including the risks and benefits of participation. The women confirmed their consent verbally. Where requested, assent was obtained from the husband or another family member in alignment with local norms. At study enrollment, the women were oriented to the developmental stage of the intervention. A total of 100 women were contacted by phone, and 41 (41%) were considered eligible based on further screening. Of these 41 women, 29 (71%) agreed to participate, and 12 (29%) declined. The 29 women were sequentially enrolled in 3 groups of 7 (24%), 10 (34%), and 12 (41%) based on their child's birth date. We sought to limit variation in birth date across groups (+2 weeks to −2 weeks) to ensure that the women were at similar postnatal stages but maintained some heterogeneity in group size for understanding group size dynamics.

Data Collection

At enrollment, sociodemographic data were collected from the participants on age, parity, village, contact number, mobile phone ownership (none, individually owned, or shared with another household member), mobile phone type (smart vs feature phone), and willingness to accept a mobile phone from the study team if they did not have their own mobile phone. Intervention feasibility and acceptability were assessed through tracking attendance, brief weekly individual check-in calls among most participants (20/29, 69%), weekly moderator reports, structured end-line in-depth interviews (IDIs) among a subset of the intervention participants (15/29, 52%), and back-end data from the mobile technologies used.

Weekly check-in calls conducted with group members were used primarily for individual problem solving and facilitative follow-up and collected data on attendance challenges, intervention acceptability (ie, level of satisfaction with the overall intervention and the intervention educational, group discussion, and SMS text messaging chat components), experience with that week's call, technical issues, perspectives on the usefulness of the educational content and group discussion, need for additional material, perspectives on changes made in intervention delivery if relevant (Table 1), and any other recommendations. In addition, after each group call, moderators noted successes, challenges, technical disruptions, significant interactions, and quality of engagement between participants and between participants and moderators and captured the participants' questions.

End-line IDIs were conducted among a subset of intervention participants (15/29, 52%) to capture their overall intervention satisfaction; their experiences with the program; the usefulness of the educational content, group discussion, and SMS text messaging group; their perspectives on participation and group dynamics; their perspectives on content and call logistics (eg, call structure, participants, and moderator preferences); and their recommendations for improving the intervention. Structured questions using a Likert-type response scale were included on overall satisfaction, usefulness of the intervention components (educational content, group discussion, and SMS text messaging group), likelihood of recommending the group, and level of connectedness.

Back-end data collected from the calls included the list of listeners, the list and time stamp of the audios played, the list of participants who *raised their hand to talk*, the number of times each participant was redialed, and the audio recording of each group call. From this, we captured the number of participants per session and the number of participants who engaged with the health worker and each other. We also gathered data from WhatsApp on SMS text messaging engagement, including posts, views, and topics. For the prerecorded audios, we collected the call duration (how long the women listened to them) and which audio recordings were accessed (topic).

Ethics Approval

Due permissions were obtained from the Indian Council of Medical Research and senior health authorities of the Government of Punjab and Mission Director, National Health Mission, India. The study protocol was approved by the University of California, San Francisco Institutional Review Board (19-299723); the Ethics Committee of the Post Graduate Institute of Medical Education and Research (IEC-03/2020-1567); the Collaborative Research Committee of the Post Graduate Institute of Medical Education and Research (79/30-Edu-13/1089-90); and the Indian Council of Medical Research (ID 2020-9576).

Results

Participant Characteristics

A total of 29 women participated in our pilot intervention (Table 3). The women ranged in age from 24 to 28 years and had all given birth within the previous 1 to 3 months for the first or second time. Educational attainment varied, with 10% (3/29) of the women being illiterate and 10% (3/29) having a postgraduate education. Infants were generally aged 3 to 4 weeks at first call. All participants (29/29, 100%) had some phone access, either a personally owned phone or a shared household phone; 59% (17/29) of the participants had a personal phone. The 41% (12/29) of participants with no personal phones were offered mobile phones and SIM cards by the research team. Of these 12 participants, 7 (58%) accepted the phones, and 5 (42%) continued to participate from shared household phones. The group members did not know each other before joining the study. Data from end-line IDI participants identified that most of the women (12/15, 80%) lived in multigenerational households with in-laws. Most end-line IDI participants (11/15, 73%) reported limited phone use and some experience with WhatsApp, Facebook, and YouTube.

Table 3. Participant characteristics, Maa Shishu Swasthya Sahayak Samooh (maternal and child health support group) pilot intervention, phase 1 (N=29).

Characteristic	Values
Age (years), median (IQR)	25 (24-28)
Parity, median (IQR)	1 (1-2)
Educational attainment, n (%)	
Illiterate	3 (10)
Primary	2 (7)
Higher primary	4 (14)
Secondary	3 (10)
Senior secondary	11 (38)
Graduation	3 (10)
Postgraduation education	3 (10)
Baby's age (weeks) at enrollment, n (%)	
1 to 2	6 (21)
3 to 4	18 (62)
5 to 6	5 (17)
Mobile network, n (%)	
Airtel	9 (31)
Idea	6 (21)
Jio	9 (31)
BSNL[a]	3 (10)
Vodafone	2 (7)
Phone owned personally, n (%)	17 (59)
Smartphone access, n (%)	28 (97)
From Punjab, n (%)	22 (76)

[a]BSNL: Bharat Sanchar Nigam Limited.

Attendance

Approximately half of the 29 participants (14/29, 48%) attended ≥3 of the 6 weekly group calls. Attendance across the 6 sessions was highest in the first few calls at 55% (16/29) for the first call and 69% (20/29) for the second call, but declined over time, with between 41% (12/29) and 48% (14/29) attending group calls 3 to 5 and 28% (8/29) attending the final weekly group call. The last group call attended was call 1 for 3% (1/29) of the participants, call 2 for 10% (3/29) of the participants, call 3 for 7% (2/29) of the participants, call 4 for 7% (2/29) of the participants, and call 5 for 34% (10/29) of the participants, and 28% (8/29) of the participants attended through the sixth group call.

Reported barriers to call attendance included infant care or other household responsibilities, visitors, network issues, and phone access. Some participants did not have their phones with them all the time, had them switched off, or participated using their husbands' phones, which were not always accessible (particularly during the workday). For example, a participant

stated that "...there was lot of work at home. My mother-in-law is sick already and with two kids, I barely get time for anything."

Intervention Acceptability

The participants reported high acceptability of the pilot intervention. Across the weekly check-ins, all of the reporting participants reached reported that they were very (16/19, 84%) or somewhat satisfied (3/19, 16%) and that they would be very likely to recommend the intervention to others (19/19, 100%). Among the end-line IDI respondents, 80% (12/15) were very satisfied, and 20% (3/15) were somewhat satisfied. Feedback on the intervention elements is detailed in the following sections.

Educational Content

Across the weekly check-ins, all of the reporting respondents (20/20, 100%) reported that the educational content was very useful. All end-line IDI respondents reported the educational content to be very useful (9/15, 60%) or somewhat useful (6/15, 40%). The participants shared their appreciation that content on both maternal and infant concerns was included and felt that the educational content met their needs. However, a couple of participants mentioned already being familiar with most of the educational content. Of the 15 end-line IDI respondents, 13 (87%) preferred receiving educational content in the form of audio recordings, mentioning that the information was always accessible and could be saved and listened to again and forwarded to others.

Group Discussion

All participants indicated that the group discussion was very useful (17/20, 85%) or somewhat useful (3/20, 15%) in the weekly check-ins. End-line IDI respondents largely found the group discussion to be very useful (13/15, 87%), with other participants reporting that it was somewhat useful (1/15, 7%) or feeling neutral (1/15, 7%). The participants appreciated the open discussion and the opportunity for everyone to ask questions. A woman shared that "everybody got time to discuss their queries and got their answers." Most participants appreciated the way the group orientation of the discussion allowed all participants to learn from the questions and experiences of their peers that were brought up on the call and believed that the time allocated for discussion was adequate. Some women suggested that the opportunity for one-to-one discussion could make women more comfortable with certain questions as "there are few topics on which we hesitate to talk in front of other people" and would be helpful for urgent queries. However, others felt that, if this were the case, that material should be brought back to the group for learning; for example, "...we can talk to each other. So [if] one-to-one talk can be done, it will be helpful. Sometimes women don't open up to doctors easily so when we talk to each other, we get to know that we are going through same thing. It will be better." The moderators noted that, although participant comfort with asking questions of the moderators improved over time, there was some reluctance to interact directly with each other with little unprompted engagement. Despite this, 79% (15/19) of the weekly check-in respondents reported feeling very or somewhat connected to other members of their group.

Text Messaging Chat

Back-end data of the WhatsApp groups showed that the number of communications ranged by group, from 59 in group 1 to 114 in group 3. Most were SMS text messages (74/114, 64.9% to 47/59, 80% across the groups), with some audio messages (8/59, 14% to 21/92, 22.8% and 26/114, 22.8% across the groups) and some images and videos (6/92, 7% to 12/114, 10.5% across the groups). The proportion of communications initiated by the participants (vs moderators) ranged from 21% (19/92) to 43% (49/114) across the groups. Educational needs were voiced in 9% (4/47) to 7.9% (9/114) of SMS text messages across the groups and included questions on both maternal (ie, pain after cesarean section, bleeding, dietary advice, and family planning) and infant (ie, spitting up, rashes, upper respiratory infections, fever after vaccination, baby massage, and crying) health concerns. Social messages were voiced in 4% (4/92) to 15.8% (18/114) of the SMS text messages, including wishing group members a happy new year and solstice.

End-line IDI participants reported infrequent chat group use but appreciated the questions asked via both SMS text message and audio message and the moderator's rapid response. The few who reported using the SMS text messaging group largely felt comfortable; however, a couple of women reported privacy and security concerns. Several women reported checking the SMS text messaging chat once per day when their husbands returned from work if they were participating in the intervention on their husbands' phones.

Call Structure

The overall structure of the calls was considered to be good by the end-line participants, with most suggesting that the calls should last approximately 30 minutes. The participants largely reported a preference for the health education content to be prerecorded so that it could be listened to on their own time before the call where the group could discuss the content.

Call Participants

A couple of the women's husbands joined the call, seemed engaged, and asked questions, although it was unclear whether this affected the other women's comfort on the call. When asked in end-line IDIs, there was some variation in perspectives regarding whether husbands or other family members should be allowed on calls, with 67% (10/15) of respondents stating that only women should be on the call and the other 33% (5/15) stating that it would be fine if the husbands joined.

Group Moderator Preferences

Although the participants were satisfied with the pilot group moderators, most participants (14/15, 93%) felt that a physician moderator would be preferred for such an intervention. Only 3% (1/29) of the participants felt that a community health worker (accredited social health activist) would be acceptable as a moderator. Of the 29 participants, 2 (7%) felt that a female physician would be preferable, but others said that it would not make a difference.

Technical Challenges

Beyond the aforementioned challenges to attendance, call challenges reported by both participants and moderators included

difficulty hearing voices and recordings. The dialing platform used for the group calls had audible calling sounds when dialing and redialing if the participants dropped off the call. These sounds caused disturbances in the ongoing live discussion and the educational audio, thus interrupting the presentation. The group calls also suffered from significant network issues, which resulted in dropped calls and unclear audio. In addition, children were sometimes heard crying in the background when the women were unmuted and asking questions.

Other Lessons Learned

Moderators noted that fostering connection between participants felt challenging and suggested that beginning the group calls in late pregnancy could help the women form relationships before the busy early postnatal period. In addition, moderators felt that including women who were both primi- and multiparous in the intervention added depth to the conversation as multiparous women could provide additional support and mentorship to primiparous women.

Design Decisions for Pilot Phase 2

The aforementioned findings and other considerations relevant to design decisions for formalizing the next iteration of the *MeSSSSage* intervention were reviewed and discussed by the research team. Each consideration is presented in Table 4 by domain (participants, intervention, and participation and engagement), including the options considered, decisions made, and summary rationale.

Table 4. Intervention design for Maa Shishu Swasthya Sahayak Samooh (maternal and child health support group) pilot intervention, phase 2.

Domain and attributes	Options	Decision	Summary rationale
Participants			
Selection criteria: parity	Primiparous only vs both primiparous and multiparous	Prioritize primiparous women but allow some multiparous women	The intervention is likely to be more impactful for first-time parents, but optimally facilitating discussion requires a large enough group. Multiparous participants may play an important role in sharing experiences and facilitating group discussion because of their experience.
Selection criteria: birth mode	Separate groups by birth mode (vaginal and cesarean) vs keeping them together	Maintain all participants together regardless of birth mode	Women with cesarean birth have unique early postpartum recovery needs; however, with our decision to start the intervention antenatally to increase group cohesiveness, further reshuffling of groups based on birth mode would be detrimental. An extra session for cesarean births could be added.
Selection criteria: allowing other individuals to participate	Limit to women only vs allowing or encouraging others to attend; consider attendee gender	Discourage but not prohibit others besides women from attending	Prioritizing privacy and confidentiality is key. Various family members attended some group sessions and perspectives were mixed; some were open to others attending; however, approximately half reported discomfort with non–group members on the calls, particularly men. Moderators felt that husband attendees asked important questions that contributed to the group discussion. An extra session that includes husbands or family members could be added.
Intervention			
Timing of recruitment and intervention	Recruit and begin intervention antenatally vs postpartum or recruit women antenatally and begin intervention in the early postpartum period	Recruit antenatally (28-32 weeks); hold 2 antenatal sessions (at approximately 32 weeks and 36 weeks) with 6 months of weekly postpartum sessions starting at 39 weeks; keep groups together regardless of birth date	Recruiting participants and initiating the groups antenatally provides the opportunity for participants to build rapport and relationships before the postpartum period, which may be more hectic. Inclusion of 2 groups antenatally allows for the promotion of health-promoting antenatal and birth practices. These benefits outweigh the disadvantages of the wider range in birth date and infant age possible within each group. A small number of participants will be expected to leave the groups because of severe maternal or neonatal complications.
Group size	12 to 20 participants per group	Target 20 participants per group	Pilot group size ranged from 7 to 12. We were initially concerned that too many people per group would overwhelm the sessions. However, attrition reduced the number of participants per session. Larger groups can accommodate any reductions in group attendance associated with antenatal recruitment and attrition.
Frequency of group calls	Weekly, twice per week, every 2 weeks, monthly	2 antenatal sessions and weekly postpartum sessions	Participants endorsed group calls twice per week; however, we anticipated feasibility concerns and eventual attrition with this intervention burden. A predictable weekly schedule routinizes the calls. We prioritized some antenatal engagement to build group familiarity before birth and meet some antenatal health education needs.
Group call length	Between 20 to 60 minutes	Target 20 minutes, allow up to 60 minutes	Previous educational information dissemination helps calls focus on group discussion. The women seemed to drop-off with longer calls.

Domain and attributes	Options	Decision	Summary rationale
Distribution of educational information	Live sharing of educational material vs dissemination of recorded content before the group call	Distribute educational material before group calls through short audio recordings; calls will include brief educational highlights	Educational material on the call and via recordings was acceptable; however, participant drop-off was higher on longer calls. Prerecorded audios allow for educational content to be accessed via an app (smartphone users) and via an IVR[a] system (feature phone users). Moving this education outside of the group discussion shortens the meetings, which may reduce barriers to full attendance and transitions within the group discussion.
Moderators	Physicians vs midlevel health professionals	Community health officers, including nurse midwives or community physicians, with some specialist support	Midlevel health professionals balance training and skills required for the moderator role and resource use, and they typically have high technological literacy. Use of these individuals as primary moderators with access to specialists for complex concerns is likely to be a scalable approach. Specialists will be featured on a few calls.
Connecting to the group call	Have network call out vs participant calls in to network or IVR	Network calls out to participants	This approach does not require the women to initiate the call, poses the lowest cost to women (no cost), and it does not exclude feature phone users. We have identified an alternative dialing platform that minimizes the disruptions observed in pilot phase 1 (ie, dialing sounds and other audio issues).
SMS text messaging chat group	All vs some groups with SMS text messaging chat group	All participants with WhatsApp-capable phones will be added to the SMS text messaging chat group.	Some women did not have their own phone and were unwilling to accept a phone from the research team. This could challenge participation in this component of the intervention for these women, and participation in the SMS text messaging chat group using a shared phone results in privacy concerns. Text chat groups will be facilitated by intervention moderators.
Phone type (overall and across groups)	Include only smartphones vs only feature phones vs both	Include both, mix groups where appropriate	Although access to smartphones is increasing rapidly across India, vulnerable women who may benefit more from intervention participation are less likely to have smartphones. Maintaining both phone types in the intervention is more complicated but more scalable.
Participation and engagement			
Building relationships and participation	Earlier recruitment; use of introductions and icebreakers; facilitation of WhatsApp group for better group cohesion	Highlight privacy of the group; incorporate icebreakers into each session; recruit earlier (during antenatal care)	Increasing group engagement is a top priority of our intervention with our strong focus on social support. Improving participation and group dynamics requires a multi-pronged approach.
Privacy	Use names vs anonymous; privacy reminders; disclose other listeners; allow for call recording or not; provide earphones	Use the women's first names and ask how they want to be addressed; do not allow participants to record discussions without permission	There is an inherent tension between respecting privacy and confidentiality and building relationships within the group. Our moderators will prioritize the comfort level of each group's participants. Where participants are interested in recording group discussions for sharing with others, this will only be allowed with permission from the full group.
Participation mechanism	Raise hand via pressing a number vs being unmuted all the time; calling on participants vs natural flow	First focus on promoting spontaneous discussion via unmuting all and then move to calling on people as needed	Ensuring that group discussions participation is easy for our participants while limiting external noise that may distract or make it hard to hear. We will integrate more structured opportunities for building group cohesion to promote comfort with discussion.

[a]IVR: interactive voice response.

Discussion

Principal Findings

Overall, *MeSSSSage* proved to be a highly acceptable approach to provide information and social support to this hard-to-reach and undersupported population—postnatal women in rural India. However, sustaining participation over time and fostering group interactions was difficult because of 2 primary challenges. The first was technological, where our preliminary mHealth platform had many issues related to network connection and audible dialing of the back-end program. The second was due to sociocultural factors related to women's comfort and cultural norms around sharing personal experiences with other women and interacting with physicians given the hierarchical social structures. These 2 challenges are common in mHealth [47] and are addressable through continuous improvement of the design of *MeSSSSage*; the goal of this phase 1 pilot was to identify such challenges.

To address technology-related challenges, we have outsourced the management of group calls to a cloud-based service provider. Given global server access, such providers are able to circumvent the reliability issues we experienced with local network providers by supplying a panel to hold internet-based calling. In our preliminary tests with this service, none of the previously occurring bothersome network and technological issues arose. We also engaged a reputed interactive voice response service provider for prerecorded educational content dissemination and developed an app for smartphone users, enabling improved access to educational content for both feature and smartphone users.

To address the second set of challenges as a result of interpersonal interactions, we are adapting our intervention model in several ways. First, instead of recruiting women immediately after giving birth, we will recruit them during late pregnancy and hold 2 group sessions prenatally. This will allow women within the same group to get to know each other before they give birth, before they are managing a newborn infant and recovering from childbirth. We will also integrate more icebreakers, get-to-know-you games, and discussion starters into each session, which should help group members get to know each other and become more comfortable interacting. Finally, although feedback on husbands' participation was mixed, with some women supporting their participation and others feeling uncomfortable with their involvement, we have decided to not allow husbands to participate to increase the comfort of all participants. Although we realize that we cannot forcibly exclude husbands given the mobile nature of the intervention and because houses may be small and phones shared, we will encourage husbands to allow the women to participate on their own given the personal nature of pre- and postnatal concerns. Other recent research on gender dynamics around phone use in India has highlighted the role of husbands as gatekeepers to women's phone use, especially when phones are shared [48]. We will carefully monitor intervention implementation to ensure that exclusion of husbands from the intervention does not lead to increases in intimate partner violence or other relationship issues and explore future opportunities for formal engagement of husbands in similar interventions. Finally, although many women said that they liked the sessions being moderated by physicians, the moderators in the pilot and other team members felt that this might hinder women's comfort talking because of the culture of respect around hierarchy [49]. We also hypothesized that part of the reason why the women said that they liked having physicians as moderators was that that was the only type of moderator they were exposed to and because of the potential for social desirability bias given that the moderators and other research team members were all from the same institution. Therefore, we will incorporate midlevel health care providers as regular intervention moderators but will invite obstetrician-gynecologists and pediatricians to attend certain sessions to respond to participant questions and provide advice. Our selection of community health officers for this role is supported by their primary roles within Ayushman Bharat, the flagship scheme of the Government of India to achieve comprehensive universal health coverage, and the National Digital Health Mission [50,51]. We will also ensure that all moderators are women.

Conclusions

The findings of this study support the *MeSSSSage* intervention's unique combination of group-based care and mHealth approaches for improving postnatal health in a geography where significant cultural and logistical barriers exist to completing the continuum of perinatal care as an innovative and promising approach for improving both maternal and neonatal health. The importance of mHealth interventions such as ours, which can successfully provide information and social support without in-person interactions, was also emphasized by the COVID-19 pandemic context in which we piloted our intervention. The revisions to overcome the primary challenges and other considerations will be implemented in the next pilot test of the *MeSSSSage* intervention. Our second intervention pilot will incorporate continued assessment of intervention feasibility and acceptability as well as preliminary assessment of effectiveness in preparation for robust effectiveness testing in subsequent research, as indicated. Ultimately, we seek to support the health and well-being of postpartum women and their infants in South Asia and beyond through the development of efficient, acceptable, and effective intervention strategies.

Acknowledgments

The authors greatly appreciate their research participants who engaged with the pilot intervention and shared their experiences and perspectives on how to improve the working intervention strategy and their intervention moderators for their excellent work implementing and iterating their intervention. The authors thank the local senior medical officer and their medical staff, including the accredited social health activists and auxiliary nurse midwives, for their support and facilitation of their research activities at

the public health facilities and the field staff for successful data collection. They would like to acknowledge Ms Deepmala Buddhija and Mr Naveen Mutiyal for providing support for data management and Ms Mata Devi for coordinating and providing administrative support between the various collaborators. The authors also appreciate the University of California, San Francisco Bixby Early Career Investigators Works in Progress seminar participants, who provided feedback on the initial drafts of this manuscript. This study was funded by the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the US National Institutes of Health (R21HD101786); the University of California**,** San Francisco Research Evaluation & Allocation Committee; and the Scheme for Promotion of Academic and Research Collaboration of the Government of India Ministry of Education (Project ID 668). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Technology development.
[DOCX File , 13 KB - formative_v6i6e34087_app1.docx]

References

1. National Family Health Survey. International Institute for Population Sciences. URL: https://www.iipsindia.ac.in/content/nfhs-project [accessed 2022-04-28]
2. Sample registration system (SRS)-special bulletin on maternal mortality in India 2016-18. Office of the Registrar General & Census Commissioner, India (ORGI). URL: https://censusindia.gov.in/nada/index.php/catalog/34781 [accessed 2022-04-29]
3. The 17 Goals. United Nations Department of Economic and Social Affairs: Sustainable Development. 2015. URL: https://sdgs.un.org/goals [accessed 2022-04-29]
4. Chmielewska B, Barratt I, Townsend R, Kalafat E, van der Meulen J, Gurol-Urganci I, et al. Effects of the COVID-19 pandemic on maternal and perinatal outcomes: a systematic review and meta-analysis. Lancet Glob Health 2021 Jun;9(6):e759-e772 [FREE Full text] [doi: 10.1016/S2214-109X(21)00079-6] [Medline: 33811827]
5. Postnatal care for mothers and newborns highlights from the World Health Organization 2013 guidelines. World Health Organization USAID Maternal and Child Health Integrated Program (MCHIP). URL: https://www.who.int/docs/default-source/mca-documents/nbh/brief-postnatal-care-for-mothers-and-newborns-highlights-from-the-who-2013-guidelines.pdf [accessed 2022-04-29]
6. Fadel SA, Ram U, Morris SK, Begum R, Shet A, Jotkar R, et al. Facility delivery, postnatal care and neonatal deaths in India: nationally-representative case-control studies. PLoS One 2015;10(10):e0140448 [FREE Full text] [doi: 10.1371/journal.pone.0140448] [Medline: 26479476]
7. Dixit P, Dwivedi L, Gupta A. Role of maternal and child health care services on postpartum contraceptive adoption in India. SAGE Open 2017 Sep 28;7(3) [FREE Full text] [doi: 10.1177/2158244017733515]
8. ACOG Committee Opinion No. 736: optimizing postpartum care. Obstet Gynecol 2018 May;131(5):e140-e150. [doi: 10.1097/AOG.0000000000002633] [Medline: 29683911]
9. Agarwal S, Curtis S, Angeles G, Speizer I, Singh K, Thomas J. Are community health workers effective in retaining women in the maternity care continuum? Evidence from India. BMJ Glob Health 2019;4(4):e001557 [FREE Full text] [doi: 10.1136/bmjgh-2019-001557] [Medline: 31406590]
10. Kosec K, Avula R, Holtemeyer B, Tyagi P, Hausladen S, Menon P. Predictors of essential health and nutrition service delivery in Bihar, India: results from household and frontline worker surveys. Glob Health Sci Pract 2015 Jun 17;3(2):255-273 [FREE Full text] [doi: 10.9745/GHSP-D-14-00144] [Medline: 26085022]
11. Syed U, Khadka N, Khan A, Wall S. Care-seeking practices in South Asia: using formative research to design program interventions to save newborn lives. J Perinatol 2008 Dec;28 Suppl 2:S9-13. [doi: 10.1038/jp.2008.165] [Medline: 19057572]
12. Adams V, Craig S, Samen A, Bhatta S. It takes more than a village: building a network of safety in Nepal's mountain communities. Matern Child Health J 2016 Dec;20(12):2424-2430. [doi: 10.1007/s10995-016-1993-1] [Medline: 27048434]
13. Lewis S, Lee A, Simkhada P. The role of husbands in maternal health and safe childbirth in rural Nepal: a qualitative study. BMC Pregnancy Childbirth 2015 Aug 04;15(1):162 [FREE Full text] [doi: 10.1186/s12884-015-0599-8] [Medline: 26239123]
14. Sudhinaraset M, Beyeler N, Barge S, Diamond-Smith N. Decision-making for delivery location and quality of care among slum-dwellers: a qualitative study in Uttar Pradesh, India. BMC Pregnancy Childbirth 2016 Jul 07;16(1):148 [FREE Full text] [doi: 10.1186/s12884-016-0942-8] [Medline: 27387024]
15. Shukla R. ICE 360 (2016) survey: household survey of Indian consumer economy and consumer environment. Research Gate. 2016. URL: https://www.researchgate.net/project/ICE-360-2016-Survey-Household-Survey-of-Indian-Consumer-Economy-and-Consumer-Environment [accessed 2022-05-23]

16. Marcolino MS, Oliveira JA, D'Agostino M, Ribeiro AL, Alkmim MB, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
17. von Salmuth V, Brennan E, Kerac M, McGrath M, Frison S, Lelijveld N. Maternal-focused interventions to improve infant growth and nutritional status in low-middle income countries: a systematic review of reviews. PLoS One 2021;16(8):e0256188 [FREE Full text] [doi: 10.1371/journal.pone.0256188] [Medline: 34407128]
18. Ashford MT, Olander EK, Rowe H, Fisher JR, Ayers S. Feasibility and acceptability of a web-based treatment with telephone support for postpartum women with anxiety: randomized controlled trial. JMIR Ment Health 2018 Apr 20;5(2):e19 [FREE Full text] [doi: 10.2196/mental.9106] [Medline: 29678804]
19. Hussain T, Smith P, Yee LM. Mobile phone-based behavioral interventions in pregnancy to promote maternal and fetal health in high-income countries: systematic review. JMIR Mhealth Uhealth 2020 May 28;8(5):e15111 [FREE Full text] [doi: 10.2196/15111] [Medline: 32463373]
20. Munro S, Hui A, Salmons V, Solomon C, Gemmell E, Torabi N, et al. SmartMom text messaging for prenatal education: a qualitative focus group study to explore Canadian women's perceptions. JMIR Public Health Surveill 2017 Mar 07;3(1):e7 [FREE Full text] [doi: 10.2196/publichealth.6949] [Medline: 28174149]
21. van den Heuvel JF, Groenhof TK, Veerbeek JH, van Solinge WW, Lely AT, Franx A, et al. eHealth as the next-generation perinatal care: an overview of the literature. J Med Internet Res 2018 Jun 05;20(6):e202 [FREE Full text] [doi: 10.2196/jmir.9262] [Medline: 29871855]
22. Jo Y, LeFevre AE, Healy K, Singh N, Alland K, Mehra S, et al. Costs and cost-effectiveness analyses of mCARE strategies for promoting care seeking of maternal and newborn health services in rural Bangladesh. PLoS One 2019;14(10):e0223004 [FREE Full text] [doi: 10.1371/journal.pone.0223004] [Medline: 31574133]
23. Mangham-Jefferies L, Pitt C, Cousens S, Mills A, Schellenberg J. Cost-effectiveness of strategies to improve the utilization and provision of maternal and newborn health care in low-income and lower-middle-income countries: a systematic review. BMC Pregnancy Childbirth 2014 Jul 22;14:243 [FREE Full text] [doi: 10.1186/1471-2393-14-243] [Medline: 25052536]
24. Dol J, Richardson B, Murphy GT, Aston M, McMillan D, Campbell-Yeo M. Impact of mobile health interventions during the perinatal period on maternal psychosocial outcomes: a systematic review. JBI Evid Synth 2020 Jan;18(1):30-55. [doi: 10.11124/JBISRIR-D-19-00191] [Medline: 31972680]
25. Zhou C, Hu H, Wang C, Zhu Z, Feng G, Xue J, et al. The effectiveness of mHealth interventions on postpartum depression: a systematic review and meta-analysis. J Telemed Telecare 2022 Mar;28(2):83-95. [doi: 10.1177/1357633X20917816] [Medline: 32306847]
26. Manandhar DS, Osrin D, Shrestha BP, Mesko N, Morrison J, Tumbahangphe KM, Members of the MIRA Makwanpur trial team. Effect of a participatory intervention with women's groups on birth outcomes in Nepal: cluster-randomised controlled trial. Lancet 2004;364(9438):970-979. [doi: 10.1016/S0140-6736(04)17021-9] [Medline: 15364188]
27. Cunningham SD, Lewis JB, Thomas JL, Grilo SA, Ickovics JR. Expect With Me: development and evaluation design for an innovative model of group prenatal care to improve perinatal outcomes. BMC Pregnancy Childbirth 2017 May 18;17(1):147 [FREE Full text] [doi: 10.1186/s12884-017-1327-3] [Medline: 28521785]
28. Cislaghi B, Heise L. Theory and practice of social norms interventions: eight common pitfalls. Global Health 2018 Aug 17;14(1):83 [FREE Full text] [doi: 10.1186/s12992-018-0398-x] [Medline: 30119638]
29. Hajimiri K, Shakibazadeh E, Mehrizi AA, Shabbidar S, Sadeghi R. The impact of general health and social support on health promoting lifestyle in the first year postpartum: the structural equation modelling. Electron Physician 2018 Jan 25;10(1):6231-6239 [FREE Full text] [doi: 10.19082/6231] [Medline: 29588825]
30. Iwata H, Mori E, Sakajo A, Aoki K, Maehara K, Tamakoshi K. Course of maternal fatigue and its associated factors during the first 6 months postpartum: a prospective cohort study. Nurs Open 2018 Apr;5(2):186-196 [FREE Full text] [doi: 10.1002/nop2.130] [Medline: 29599994]
31. Negron R, Martin A, Almog M, Balbierz A, Howell EA. Social support during the postpartum period: mothers' views on needs, expectations, and mobilization of support. Matern Child Health J 2013 May;17(4):616-623 [FREE Full text] [doi: 10.1007/s10995-012-1037-4] [Medline: 22581378]
32. Shrestha S, Adachi K, Petrini MA, Shrestha S. Factors associated with post-natal anxiety among primiparous mothers in Nepal. Int Nurs Rev 2014 Sep;61(3):427-434. [doi: 10.1111/inr.12118] [Medline: 25039801]
33. Grassley JS. Adolescent mothers' breastfeeding social support needs. J Obstet Gynecol Neonatal Nurs 2010;39(6):713-722. [doi: 10.1111/j.1552-6909.2010.01181.x] [Medline: 20880051]
34. Reid KM, Taylor MG. Social support, stress, and maternal postpartum depression: a comparison of supportive relationships. Soc Sci Res 2015 Nov;54:246-262. [doi: 10.1016/j.ssresearch.2015.08.009] [Medline: 26463547]
35. Garcia ER, Yim IS. A systematic review of concepts related to women's empowerment in the perinatal period and their associations with perinatal depressive symptoms and premature birth. BMC Pregnancy Childbirth 2017 Nov 08;17(Suppl 2):347 [FREE Full text] [doi: 10.1186/s12884-017-1495-1] [Medline: 29143635]
36. Hale N, Picklesimer AH, Billings DL, Covington-Kolb S. The impact of Centering Pregnancy Group Prenatal Care on postpartum family planning. Am J Obstet Gynecol 2014 Jan;210(1):50.e1-50.e7. [doi: 10.1016/j.ajog.2013.09.001] [Medline: 24018309]

37. Shakespear K, Waite PJ, Gast J. A comparison of health behaviors of women in centering pregnancy and traditional prenatal care. Matern Child Health J 2010 Mar;14(2):202-208. [doi: 10.1007/s10995-009-0448-3] [Medline: 19184385]
38. Carlson NS, Lowe NK. CenteringPregnancy: a new approach in prenatal care. MCN Am J Matern Child Nurs 2006;31(4):218-223. [doi: 10.1097/00005721-200607000-00004] [Medline: 16940816]
39. Eluwa GI, Adebajo SB, Torpey K, Shittu O, Abdu-Aguye S, Pearlman D, et al. The effects of centering pregnancy on maternal and fetal outcomes in northern Nigeria; a prospective cohort analysis. BMC Pregnancy Childbirth 2018 May 11;18(1):158 [FREE Full text] [doi: 10.1186/s12884-018-1805-2] [Medline: 29751797]
40. Trotman G, Chhatre G, Darolia R, Tefera E, Damle L, Gomez-Lobo V. The effect of centering pregnancy versus traditional prenatal care models on improved adolescent health behaviors in the perinatal period. J Pediatr Adolesc Gynecol 2015 Oct;28(5):395-401. [doi: 10.1016/j.jpag.2014.12.003] [Medline: 26233287]
41. Selvaraj R, Ramakrishnan J, Sahu SK, Kar SS, Laksham KB, Premarajan KC, et al. High prevalence of anemia among postnatal mothers in Urban Puducherry: a community-based study. J Family Med Prim Care 2019 Aug;8(8):2703-2707 [FREE Full text] [doi: 10.4103/jfmpc.jfmpc_386_19] [Medline: 31548960]
42. National Family Health Survey - 4 2015 -16: State Fact Sheet Punjab. International Institute for Population Sciences. URL: http://rchiips.org/nfhs/pdf/NFHS4/PB_FactSheet.pdf [accessed 2022-04-29]
43. Sidhu TK, Coonar PP. Contraceptive usage and awareness among postpartum mothers in urban field practice area of a tertiary hospital. Indian J Community Health 2015;27(1):139-142 [FREE Full text]
44. Upadhyay RP, Chowdhury R, Salehi A, Sarkar K, Singh SK, Sinha B, et al. Postpartum depression in India: a systematic review and meta-analysis. Bull World Health Organ 2017 Oct 01;95(10):706-17C [FREE Full text] [doi: 10.2471/BLT.17.192237] [Medline: 29147043]
45. Michie S, Johnston M, Francis J, Hardeman W, Eccles M. From theory to intervention: mapping theoretically derived behavioural determinants to behaviour change techniques. Applied Psychol Int Rev-Psychologie Appliquee-Revue Internationale 2008 Oct;57(4):660-680. [doi: 10.1111/j.1464-0597.2008.00341.x]
46. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
47. Gurupur VP, Wan TT. Challenges in implementing mHealth interventions: a technical perspective. Mhealth 2017;3:32 [FREE Full text] [doi: 10.21037/mhealth.2017.07.05] [Medline: 28894742]
48. Hersh S, Nair D, Komaragiri PB, Adlakha RK. Patchy signals: capturing women's voices in mobile phone surveys of rural India. BMJ Glob Health 2021 Sep;6(Suppl 5):e005411 [FREE Full text] [doi: 10.1136/bmjgh-2021-005411] [Medline: 34475116]
49. Claramita M, Nugraheni MD, van Dalen J, van der Vleuten C. Doctor-patient communication in Southeast Asia: a different culture? Adv Health Sci Educ Theory Pract 2013 Mar 8;18(1):15-31 [FREE Full text] [doi: 10.1007/s10459-012-9352-5] [Medline: 22314942]
50. Induction training module for community health officers. National Health Systems Resource Centre. URL: http://www.nrhmorissa.gov.in/writereaddata/Upload/Documents/Induction%20Training%20Module%20for%20CHOs.pdf [accessed 2022-05-18]
51. Desai S, Bishnoi RK, Punjot P. Community health officer: the concept of mid-level health care providers. Int J Community Med Public Health 2020 Mar 26;7(4):1610. [doi: 10.18203/2394-6040.ijcmph20201483]

Abbreviations

ANM: auxiliary nurse midwife
IDI: in-depth interview
MeSSSSage: Maa Shishu Swasthya Sahayak Samooh (maternal and child health support group)
mHealth: mobile health

Edited by A Mavragani; submitted 09.10.21; peer-reviewed by J Dol, N Slabyj; comments to author 17.01.22; revised version received 12.04.22; accepted 18.04.22; published 29.06.22.

Please cite as:
El Ayadi AM, Duggal M, Bagga R, Singh P, Kumar V, Ahuja A, Kankaria A, Hosapatna Basavarajappa D, Kaur J, Sharma P, Gupta S, Pendse RS, Weil L, Swendeman D, Diamond-Smith NG
A Mobile Education and Social Support Group Intervention for Improving Postpartum Health in Northern India: Development and Usability Study
JMIR Form Res 2022;6(6):e34087
URL: https://formative.jmir.org/2022/6/e34087
doi:10.2196/34087
PMID:35767348

Original Paper

Possible Contribution of Meaning in Life in Patients With Chronic Pain and Suicidal Ideation: Observational Study

Vasileios Chytas[1*], MD; Alessandra Costanza[2,3*], MD; Viridiana Mazzola[1], PhD; Christophe Luthy[4,5], MD, Prof Dr; Vasiliki Galani[1], MD; Guido Bondolfi[1,2], MD, Prof Dr; Christine Cedraschi[4,5,6], Prof Dr

[1]Department of Psychiatry, Service of Liaison Psychiatry and Crisis Intervention, Geneva University Hospitals, Geneva, Switzerland

[2]Department of Psychiatry, Faculty of Medicine, University of Geneva, Geneva, Switzerland

[3]Faculty of Biomedical Sciences, Università della Svizzera Italiana, Lugano, Switzerland

[4]Division of General Medical Rehabilitation, Geneva University Hospitals, Geneva, Switzerland

[5]Department of Geriatrics and Rehabilitation, Faculty of Medicine, University of Geneva, Geneva, Switzerland

[6]Division of Clinical Pharmacology and Toxicology, Multidisciplinary Pain Centre, Geneva University Hospitals, Geneva, Switzerland

[*]these authors contributed equally

Corresponding Author:
Vasileios Chytas, MD
Department of Psychiatry
Service of Liaison Psychiatry and Crisis Intervention
Geneva University Hospitals
Rue Gabrielle-Perret-Gentil 4
Geneva, 1205
Switzerland
Phone: 41 079 553 41 55
Email: Vasileios.Chytas@hcuge.ch

Abstract

Background: Chronic pain is associated with an elevated risk of suicidal ideation (SI).

Objective: We aimed to examine if the presence or the search for Meaning in Life (MiL) are associated with less SI and explore whether MiL profiles emerge in our cohort. These profiles can be described as high presence–high search, high presence–low search, low presence–low search, and low presence–high search.

Methods: In this observational study, we recruited 70 patients who were referred to the Multidisciplinary Pain Center of the Geneva University Hospitals and who answered positively to question 9 on the Beck Depression Inventory, 2nd Edition, investigating SI. Patients who agreed to participate in the study were further investigated; they participated in a structured diagnostic interview to screen for psychiatric diagnoses. During this interview, they completed the Meaning in Life Questionnaire and the semistructured Scale for Suicide Ideation (SSI) to assess the characteristics and severity of SI.

Results: There was a statistically significant correlation between the presence of MiL subscale and the SSI. These 2 scales had a negative and statistically highly significant correlation (R=–.667; P<.001). The results also showed a negative and statistically highly significant correlation between the score of the search for MiL and the SSI (R=–.456; P<.001). The results thus pointed to the presence of MiL as a potential protective factor against the severity of SI, while the search for MiL is also a possible resiliency factor, although to a lesser extent. The profile low presence–low search grouped the vast majority (47%) of the patients; in these patients, the mean SSI score was 14.36 (SD 5.86), much higher compared with that of the other subgroups.

Conclusions: This study's results point to MiL as a concept of interest regarding devising psychotherapeutic interventions for chronic pain patients in order to reduce the suicidal risk and more accurately determine patients' suffering.

(JMIR Form Res 2022;6(6):e35194) doi:10.2196/35194

KEYWORDS

meaning in life; suicidal ideation; chronic pain; pain; suicide

Introduction

Patients suffering from chronic pain conditions are at an elevated risk of suicidal behavior (SB). The literature points to a 20% to 40% prevalence rate of suicidal ideation (SI); lifetime prevalence of suicidal attempts (SA) is estimated between 5% and 14%; further, the risk of completed suicide is doubled in patients with chronic pain as compared to controls [1-4]. In most studies, the associations between chronic pain and SB were robust even after adjusting for the effect of sociodemographic characteristics and psychiatric comorbidities, including depressive conditions [5-11]. Numerous specific conditions that can modulate the SB risk in patients with chronic pain have been investigated (eg, pain characteristics, functional interference, illness beliefs, access to opioids) [1,3,12].

However, the literature also highlights the need for further exploration of other possible risk and protection factors to improve the characterization of the SB risk profile in these patients [3]. Using this framework, we will explore the role of meaning in life (MiL) as a resiliency factor possibly modulating SI in patients with chronic pain. Experiencing chronic pain or any other illness often requires revising one's life goals and expectations [13]. Hence, its effect is not only limited to the patients' biopsychosocial functioning, but it also affects the existential domain [14]. This common clinical observation raises the question of MiL in chronic pain patients with SI. A reduction of SI in patients attending an interdisciplinary treatment for chronic pain has been reported in approximately one-third of the participants [15]. A larger investigation of SB risk and resiliency factors in chronic pain, beyond the well-known demographic and psychiatric factors, has been strongly advocated for [3].

According to the conceptualization of MiL, this construct is a web of connections that help to read and understand the pain experience and create strategies directed at achieving the expected outcomes regarding reducing pain and its consequences [16]. This model divides MiL into 2 constructs, presence of and search for MiL [17], which are not mutually exclusive [18,19]. Many consider the presence of MiL as beneficial [20]; the search for MiL appears more controversial with some authors considering it the essence of human motivation [21] and others as a sign that one's life has lost meaning [22] or has less meaning [23,24]. Different MiL profiles recently have been characterized in patients with chronic pain, resulting from the combination between low and high levels of presence and search for MiL, which were associated with a unique adjustment outcome: patients having profiles with high scores of presence showed fewer depressive symptoms and greater life satisfaction [14,25]. The literature does not yet mention the association of MiL with SI in patients with chronic pain.

Drawing on these studies and the MiL construct, this study aimed to investigate the relationship between MiL and intensity of SI in patients with chronic pain presenting SI.

Our objectives were as follows:

- Examine whether the presence of or search for MiL are associated with less SI
- Explore whether previously described MiL profiles (high presence–high search, high presence–low search, low presence–low search, and low presence–high search) emerge in our cohort

This study's results will permit future exploration into further psychotherapeutic interventions integrating MiL profiles on the model of already existent standards for patients with other somatic conditions [26-29].

Methods

Setting and Procedure

The Division of Clinical Pharmacology and Toxicology and the Service of Liaison Psychiatry and Crisis Intervention at the Multidisciplinary Pain Center (MPC) of the Geneva University Hospitals conducted this observational study [30]. The MPC is a third-line ambulatory referral center to which treating physicians refer most patients for an interdisciplinary clinical evaluation and review of treatment proposals (eg, physical treatment, individual or group psychiatric/psychological treatment, or pharmacological proposals).

Participants were enlisted via ongoing recruitment through a project anchored in daily clinical practice at the MPC. Each participant received a series of self-administered screening questionnaires at home before the first routine visit to the MPC. All patients received these questionnaires regardless of their participation in this study because these questionnaires are routinely used at the MPC and they constitute an important clinical and professional tool. These questionnaires included the Beck Depression Inventory, 2nd Edition (BDI-II) [31]. In a second step, approximately 15 days after receiving the self-administered routine screening questionnaires, patients were seen at the MPC. During this consultation, patients with SI (ie, those who answered positively [responses 1, 2, and 3] to question 9 of the BDI-II) received written and oral information about this study from a member of the medical team who also responded to their questions. If serious suicidal thoughts were identified, all appropriate measures were taken, including, if necessary, accompanying patient to the psychiatric emergency ward at the Geneva University Hospitals. In a third and final step, 1 to 7 days after the visit to the MPC, patients with SI who agreed to participate in the study attended a second visit. During this consultation, patients signed informed consent and were included in the study. They also participated in a clinical interview and structured diagnostic interview, the French version 5.0.0 of the Mini-International Neuropsychiatric Interview (MINI) [32], to screen for psychiatric diagnoses according to the *Diagnostic and Statistical Manual of Mental Disorders, 4th Edition* (DSM-IV). During this same interview, patients completed the Meaning in Life Questionnaire (MiLQ) [23]. In addition, a clinical evaluation of SI was conducted as well as a semistructured scale, the Scale for Suicide Ideation (SSI), to assess the characteristics and severity of SI [33].

Inclusion criteria were patients with chronic pain, referred to the MPC, presenting with an SI, indicated by a positive response (responses 1 to 3) to question 9 of the BDI-II, aged over 18 years, and providing written informed consent. Exclusion criteria

were insufficient understanding of the French language, diagnosis of organic mental disorders (F00-F09), psychotic disorders (F20-F29), or borderline personality disorder (F60.3).

Instruments

We used the instruments presented in Table 1.

Table 1. Instruments used for the assessment.

Instruments	Main characteristics of the instruments
Question 9 of the Beck Depression Inventory [31]	• Self-report multiple choice inventory • Indicator of the severity of depressive symptoms; standard cutoff scores: 0-10 (no depression), 11-19 (mild depression), 20-29 (moderate depression), >30 (severe depression) • 21-items, each rated on a 4-point scale ranging from 0-3 based on severity of the item • Question 9: "Suicidal thoughts or wishes" (0: I don't have any thoughts of killing myself; 1: I have thoughts of killing myself, but I would not carry them out; 2: I would like to kill myself; 3: I would kill myself if I had the chance).
Scale for Suicide Ideation [33,34]	• Scale based on a semistructured interview with the patient • Indicator of characteristics/severity of an individual's plans and wishes to commit suicide • 19 items, each rated on a 3-point scale ranging from 0-3 based on severity of the item • Total for the 19 items: minimum=0, maximum=38 (higher scores indicate greater SI[a]); score ≥6 indicates clinically significant SI [35,36]
Meaning in Life Questionnaire [23]	• Measure of presence of and search for (5 questions each) MiL[b] • For each subscale, scores range from 5-35, high scores indicating high presence of or search for MiL. MiLQ[c] does not have definitive cutoff scores as it measures MiL across the range of human functioning; the author provides probabilistic estimates about scores above, equal, or below 24 on presence of and search for constructs [37].

[a]SI: suicidal ideation.

[b]MiL: meaning in life.

[c]MiLQ: Meaning in Life Questionnaire.

The BDI-II [31] has been repeatedly used in the context of chronic pain [38,39] and has undergone extensive validation, including in the French language [40,41]. The importance of somatic symptoms of depression in pain patients has been stressed along with the risk of a score inflation in people with chronic pain [42] and hence the importance of using a structured clinician-administered interview to reliably identify a possible depressive disorder [43].

Considering the BDI-II screens depressive symptoms and is not a diagnostic tool, an experienced psychiatrist also conducted an interview using the MINI, which is extensively used to screen for probable psychiatric diagnoses according to the DSM-IV and ascertain the presence of depression in various somatic or psychiatric contexts. The French version has been used [44].

The SSI [33,34] assesses the characteristics and severity of SI. This interviewer-administered scale is one of the most widely used instruments for assessing suicidal thinking. It helps to identify suicidal individuals if they are willing to acknowledge and share their thoughts. The SSI serves as a routine screening for existent suicidal thinking and can aid in an exploration that is more extensive regarding the severity of such thoughts. It can be administered in various settings (eg, general medical services, psychiatric services) and during a routine screening [33]. The SSI has a validated French version [45,46] that has already been used in the context of the Geneva University Hospitals in Switzerland [47].

The MiLQ [23] has 2 subscales: presence of and search for MiL. The MiLQ has been translated into French [47] and used in the context of the Geneva University Hospitals in Switzerland [47,48]. Other researchers have used it with patients suffering from chronic pain [25], depression [49], or other chronic health problems [14]. We further used the MiLQ to define the profiles of MiL (high presence–high search, high presence–low search, low presence–low search, and low presence–search high) as other authors suggested and used the same method of analysis on the results obtained in our sample [14].

Statistical Analyses

Data from all questionnaires were manually entered into an SPSS database (IBM Corp). Double data entry was performed to minimize entry errors. Questionnaires and demographic data were analyzed with SPSS 22. Descriptive statistical analyses were performed to calculate the means and standard deviations of the numeric variables and the frequencies and percentages of the categorical variables. The chi-square test was used for categorical variables and to check for correspondence between groups with respect to age, sex, and years of education. To test our main hypothesis, a nonparametric statistical dependence measure (Spearman rho) or a Pearson correlation coefficient was calculated to look for any significant interaction between the MiLQ subscales (ie, presence of or search for) and SI. As for missing values, the analysis was performed across all questionnaire scores to look for significant trends. Random missing values were replaced with the mean.

Ethics Approval

The Ethics Committee of the Canton of Geneva approved the protocol (project No. 2017-02138; decision dated January 25, 2018) that has been carried out in accordance with the research plan and Swiss legal and regulatory requirements, which are in

agreement with the principles stated in the current version of the Declaration of Helsinki. We obtained written informed consent from all patients.

Results

Participants

A total of 70 patients with SI (ie, patients who answered >1 to question 9 of the BDI-II) were recruited. These patients were enlisted for 20 months between March 2018 and November 2019; 8 patients did not wish to participate in the study and another 4 were excluded because they did not have a sufficient mastery of the French language. These 82 patients who were initially contacted to participate in our study consist of about 15% of the total number of patients who were referred to the MPC during this period.

Table 2 presents the sociodemographic and clinical characteristics of the 70 patients who agreed to participate in the study. Of these, 60% (42/70) were women, had a mean age of 54 years, and were mostly professionally qualified. Only 14% (10/70) of patients were currently employed; 77% (54/70) were on sick leave due to pain, partial or full time, and among those, only 15% (11/54) of patients were engaged in professional reintegration measures.

Clinical Characteristics

The origin of the pain was neuropathic or nociceptive in most of the patients. They had a pain history that was of long duration, with an average of 8 years, but also of high intensity, with an average intensity of current pain close to 7 out of 10 on a 100-mm visual analog scale (VAS), and as many as one-third of the patients (23/70, 32.9%) assessing the maximum intensity of their pain at 10 out of 10.

Table 2. Sociodemographic and clinical characteristics of the participants.

	Values
Gender, n (%)	
Male	28 (40)
Female	42 (60)
Age (years), mean (SD)	54.26 (14.5)
Civil status, n (%)	
Married	31 (44)
Single	13 (19)
Separated/divorced	21 (30)
Widowed	5 (7)
French as original language, n (%)	42 (60)
Education, n (%)	
Primary school	13 (19)
Apprenticeship	27 (38)
Vocational school	21 (30)
University	9 (13)
Present professional activity, n (%)	10 (14)
Pain duration (years), mean (SD)	8.1 (8.1)
Pain intensity, mean (SD)	
VAS[a] current	6.9 (2.4)
VAS minimum	5.3 (2.8)
VAS maximum	8.9 (1.4)
Origin of pain, n (%)	
Neuropathic	44 (63)
Nociceptive	22 (31)
Visceral or other	4 (6)

[a]VAS: visual analog scale.

Most of the participants considered their pain as related to a specific event. This event was an accident in 30% (21/70), one or several surgeries or treatments were perceived as harmful in 15% (11/70), and emotional factors and traumatic life events occurred in another 12% (9/70). Some (17/70, 25%) of the patients attributed their pain to a disease and 10% (7/70) to

heavy labor. Only about 8% (5/70) of the patients could not think of a specific cause.

Depressive Symptomatology

The mean score of the BDI-II was 31.26 (SD 11.37). More precisely, the absence of depressive symptoms (BDI-II ≤10) was rare in our sample (2/70, 3%). On the other end of the spectrum, as many as 49% (34/70) of patients had a BDI-II score over 30, indicating symptomatology present during a severe depressive episode. Regarding the other patients, 14% (10/70) had a score between 11 and 20 (corresponding to a mild depressive episode) and 34% (24/70) had a score between 21 and 30 (depressive symptoms compatible with a moderate depressive episode).

The MINI structured clinical interview confirmed the presence of a depressive episode in the majority of these patients (68/70, 97%) and highlighted other psychiatric comorbidities among them, in particular anxiety disorders (panic disorder, social phobia, agoraphobia, and obsessive and compulsive disorder), which were largely represented (25/70, 36%).

MiL and SI

The mean score on the SSI was 11.40 (SD 5.92). As a reminder, high scores on this scale indicate high SI, and the score of 6 is used as a cutoff to indicate clinically significant SI. In this group, 83% (58/70) of patients had an SSI score ≥6 and 17% (12/70) had a score of <6.

In our group, there was no significant correlation between the presence of or search for MiL and the intensity or duration of pain. Gender, origin of pain, marital status, or age also seemed not to affect the presence of or search for MiL. However, we found a significant difference in the presence of MiL between working and nonworking patients (P=.04) but not in the search for MiL.

We also investigated the relationship between MiL and SI, considering separately the 2 concepts, presence of and search for MiL. For each subscale, the scores vary between 5 and 35, with high scores signifying a high presence of or search for MiL. The cutoff for these 2 subscales is 24, with scores equal or higher indicating higher presence of or search of MiL. Regarding the presence of MiL, the average score on the presence of MiL subscale was 20.13 (SD 8.23). Among the patients in the study, 37% (26/70) presented a score for the presence of MiL ≥24; for 63% (44/70) the score was <24. Regarding the search for MiL, the average score on the subscale was 18.14 (SD 8.64); 30% (21/70) of patients had a score ≥24, and 70% (49/70) had a score of <24. There was also a positive and significant correlation between the presence of and the search for MiL (R=.402; P=.001).

Furthermore, there was a statistically significant correlation between the presence of MiL subscale and the SSI. These 2 scales had a negative and statistically highly significant correlation (R=–.667; P<.001). The results also showed a negative and statistically highly significant correlation between the score of the search for MiL and the SSI (R=–.456; P<.001).

Considering the results of the MiLQ in the 2 subgroups (ie, patients who presented a high SI [SSI ≥6] and those presenting lower SI [SSI <6]) highlighted the following results: in the first subgroup (SSI ≥6), the negative correlation between the severity of SI and the presence of MiL was statistically very significant (R=–.544; P<.001). This group also had an average score on the MiL presence subscale (18.19 [SD 7.35]) that was lower than the average score of the whole group (20.13). In contrast, in patients with a lower score on the SSI (SSI<6), the correlation between the presence of MiL and SI was negative but not significant and the mean score on the MiL presence subscale, considerably higher, was 29.50 (SD 5.44), above the mean score for the whole group.

The results were more contrasted for the search for MiL in these 2 subgroups. In the subgroup of patients with higher SI (SSI≥6), the mean score for the search for MiL was 17.52 (SD 8.84), close to the mean score of the whole group (18.14). In this subgroup, there was a strong and significant negative correlation between the search for MiL and SI (R=–.492; P<.001). In the subgroup of patients with a weaker SI (SSI<6), the mean score for the search for MiL was 21.17 (SD 7.17), higher than the mean score of the whole group (18.14). However, in this subgroup, the correlation between the search for MiL and SI was not significant.

Finally, we found a statistically significant difference concerning the presence of MiL subscale's score between the 2 subgroups, patients with higher SI (SSI≥6) and those with lower SI (SSI<6), with P<.001. However, regarding the subscale of the search for MiL, the difference between these 2 subgroups was not significant.

Profiles of MiL

We examined the relationship of the different profiles of MiL with SI, using the grouping Dezutter et al [14] described (ie, applying the same analysis to the scores obtained in our sample). In our study, 23% (16/70) of patients corresponded to the profile high presence–low search; those patients had an average score on the SSI of 8.56 (SD 5.11). On the other end, 16% (11/70) of patients who corresponded to the profile low presence–high search had an average score on the SSI of 10.45 (SD 4.30). The profile high presence–high search grouped 14% (10/70) of patients with an average SSI score of 7.20 (SD 3.99). The profile low presence–low search grouped the majority of the patients (33/70, 47%); in these patients, the mean SSI score was 14.36 (SD 5.86), which was much higher than that of the other subgroups.

Discussion

Principal Findings

This study showed that MiL is associated with SI in patients with chronic pain. Our results showed that the presence of MiL seemed to be a potentially protective factor against SI. The statistically significant negative association between the subscale presence of MiL and SSI, with the presence of MiL actually associated with a decrease in the severity of SI, further supports our results. We found this negative correlation both in the patients with a high SI (SSI ≥6) and in those who presented a lower SI (SSI <6). These results are in line with those of the literature, which highlight this negative correlation between the

presence of MiL and SI among students [50], military and veterans, [28] and HIV-positive patients [51]. These data thus support our initial hypothesis, which held that the more a patient identifies meaning in their life, the lower the severity of their SI.

The results are less straightforward regarding the search for MiL. In our sample, we found a negative and significant correlation between the search for MiL and SI. This means that the patients who seek MiL, regardless of the presence of MiL, have less intense SI as compared to those who do not. Furthermore, this significant negative correlation is also found in the subgroup of patients who present with high SI (SSI ≥6), but this association is not significant in patients with less intense SI (SSI <6).

Contrary to the results for the presence of MiL, the difference in the scores on the search for MiL subscale is not significant between patients with strong SI (SSI ≥6) and patients with less intense SI (SSI <6). This shows that while the presence of MiL is clearly higher in patients with lower SI, the search for MiL tends not to be significantly different between patients with high versus low SI. Previous studies have also found a protective effect of searching for MiL against SI [52], but other studies have reported a positive correlation between searching for MiL and SA [28]. Overall, our results support the search for MiL as a factor that can protect against an increased severity of SI; however, less significantly as compared to presence of MiL.

In our group, neither presence of nor search for MiL were associated with sociodemographic or pain-related characteristics, except for professional activity, which was associated with the presence of MiL. This is also in line with previous research suggesting that patients who present MiL display a higher level of adaptation, including in the professional field [24]. Fewer working individuals seeking disability may mediate these results, as well as less MiL related to the professional sphere in these participants. This issue may be worth addressing in future studies.

Four profiles of MiL (high presence–low search, high presence–high search, low presence–high search, and low presence–low search) have been described in chronic pain patients [14]. The patients in the high presence–high search for MiL subgroup had a much lower SSI score compared to the other subgroups, followed by the high presence–low search subgroup.

The literature has reported that when the search for MiL is high and associated with a high presence of MiL, the determining effect of the search for meaning appears to decrease [14]. Therefore, in these patients, we believe that because they consider their life as having meaning, the additional search for meaning represents a healthy approach and that they thus present a good level of well-being and acceptance [14]. On the other hand, what differentiates our study from that of the Dezutter et al study [14] is that the first subgroup, high presence–high search for MiL, has fewer suicidal thoughts compared to the second, high presence–low search. Our results suggest that the search for meaning also plays its probable protective part against SI. This may be at least partly related to these patients' expectations of the pain consultation as a means to relieve their pain and improve their quality of life while simultaneously seeking to give them meaning. Therefore, we see in our study in a more pronounced way an increased protective role of research as a lever of vital impetus in the patients who contact our center.

However, the search for MiL when no meaning is presently identified is associated with higher SI scores, thus pointing to the search for meaning as more anxiety provoking and stressful. Some have suggested that such a pattern might result in an adaption that is more problematic and less optimal psychological well-being [14]. In our study, we also identified patients corresponding to the profile high search–low presence of MiL and presenting with a higher average score on the SI scale when compared to the first 2 subgroups. Furthermore, we found a significantly higher SI score in patients who neither present nor search for an MiL. These patients constitute the group with the lowest level of adaptation and lowest well-being in the Dezutter et al [14] study, which, as we have done, examined the link between chronic pain and MiL. However, these results contrast with those of Steger et al [18], which suggests that for patients who fail to identify meaning in their life, seeking a meaning leads to a level of adaptation and to a particularly low well-being, even more than patients who neither present nor search for a meaning. In this case, the research shows it having an anxiety-inducing, even deleterious role. In the context of varying adaptation levels, the issue of psychological flexibility is of interest. Psychological flexibility has indeed been described as a key feature in the psychological approaches to chronic pain management [53] and has been largely used in the protocol called acceptance and commitment therapy (ACT). ACT treatment programs have received attention not only in the field of chronic pain but also as a means of reducing SI [54]. Hence, considering MiL and reducing SI as possible targets of ACT for chronic pain warrants further investigation.

Finally, our data confirmed that the presence of and the search for MiL are positively and significantly correlated, which suggests that the 2 constructs may depend on one another. This is not in line with previous research [23,24,47], but the role of intervening variables, acting as mediating factors, has not been explored in this study and clearly warrants further investigation.

Limitations

Our study has limitations that need acknowledgment. First, the study sample is small, and thus not allowing for the construction of subgroups using the participants' sociodemographic and clinical characteristics; however, the sample allowed investigating and categorizing the patients according to the MiL and SI dimensions. Second, the MiLQ has not been extensively validated in French; however, it has been used in various contexts, particularly psychiatric emergencies, in its current version [47]. Third, and importantly, the participants recruited in this study were only those who self-reported (ie, admitted) SI, and yet admitting SI has been described as highly problematic [55,56]. Indeed, both underreporting as a means of concealing symptoms [3,57] and overreporting as a means of soliciting increased attention or making a plea for help [58] may induce response bias. Patients who underreported SI may be underrepresented while those who may be prone to overreport may in turn be overrepresented, thus limiting the findings'

generalizability. Self-report screening measures have inherent limitations regarding eliciting the individuals' phenomenological experiences but also knowing whether they provide an accurate report of the clinical phenomenon under investigation [59,60]. Fourth, this is a cross-sectional study, and this design does thus not allow for any inferences regarding a causal relationship between changes in MiL as leading to a reduction in SI. Taken together, these limitations also point to the existence of biases and possible confounders as one of the key methodological issues of correlational studies [61].

Conclusion

Taken together, our results regarding the different MiL profiles and associated risk factors highlight the place to be given to individual participants. Indeed, MiL profiles may allow clinicians to consider mental and somatic suffering; however, the search for the coherence of a constellation (eg, pain status, functional interference, illness beliefs, emotional context, SI, and SA) should not mask the specific form of the individual patient's suffering. Using MiL profiles can help clinicians recognize their patient's suffering while keeping in mind the necessity to consider the patient's singularity even while using validated tools to describe mental states and organization. In chronic pain patients, the continuity and attention of clinicians to maintain the associative work of the patient's discourse is of main concern; indeed, facing suicidal crisis, the process at work in the patients can hinder this work. Considering the concept of MiL should support clinical work aimed at providing the most adequate response to patient suffering, as it has proven possible in the context of psychiatric emergencies [47,62]. Psychotherapeutic interventions targeting MiL have been found effective in reducing suicide risk [26] and represent a promising therapeutic opportunity [27-29] in various clinical contexts. However, this construct's contribution to decreasing suicide risk in patients with chronic pain remains to be established. Our study may help find further tools to treat suffering and emotional distress in chronic pain patients through defining a unique and individualized suicide risk profile, improving screening and preventive strategies, and developing intervention strategies including the existential sphere. Clinical intervention research would be beneficial in determining if interventions based on active bolstering of MiL, such as methods of linking listening to the patients and understanding their phenomenological experience with strategies for increasing MiL, can affect outcomes for individuals presenting with SI and SB. The Attempted Suicide Short Intervention Protocol is such a treatment method to consider for exploring the patient's experience and the follow-up [63-66].

Our study supports a probable protective role of MiL against the severity of SI in patients with chronic pain presenting with SI. Further studies should be conducted to better enlighten the complex relationship between suicidal risk, prevention, and intervention strategies in chronic pain patients to provide new perspectives. The results stress the importance of thorough patient interviews regarding bringing theoretical conceptualizations into daily clinical practice.

Acknowledgments

We are grateful to all the patients who gave their valuable time and agreed to participate in our study.

Conflicts of Interest

None declared.

References

1. Tang NKY, Crane C. Suicidality in chronic pain: a review of the prevalence, risk factors and psychological links. Psychol Med 2006 May;36(5):575-586. [doi: 10.1017/S0033291705006859] [Medline: 16420727]
2. Fishbain DA, Lewis JE, Gao J. The pain suicidality association: a narrative review. Pain Med 2014 Nov 01;15(11):1835-1849. [doi: 10.1111/pme.12463] [Medline: 24995953]
3. Hassett AL, Aquino JK, Ilgen MA. The risk of suicide mortality in chronic pain patients. Curr Pain Headache Rep 2014 Jun 22;18(8):436. [doi: 10.1007/s11916-014-0436-1] [Medline: 24952608]
4. Calati R, Laglaoui Bakhiyi C, Artero S, Ilgen M, Courtet P. The impact of physical pain on suicidal thoughts and behaviors: meta-analyses. J Psychiatr Res 2015 Dec;71:16-32. [doi: 10.1016/j.jpsychires.2015.09.004] [Medline: 26522868]
5. Ilgen MA, Kleinberg F, Ignacio RV, Bohnert ASB, Valenstein M, McCarthy JF, et al. Noncancer pain conditions and risk of suicide. JAMA Psychiatry 2013 Jul 01;70(7):692-697. [doi: 10.1001/jamapsychiatry.2013.908] [Medline: 23699975]
6. Kikuchi N, Ohmori-Matsuda K, Shimazu T, Sone T, Kakizaki M, Nakaya N, et al. Pain and risk of completed suicide in Japanese men: a population-based cohort study in Japan (Ohsaki Cohort Study). J Pain Symptom Manage 2009 Mar;37(3):316-324 [FREE Full text] [doi: 10.1016/j.jpainsymman.2008.03.012] [Medline: 18694632]
7. Ratcliffe G, Enns M, Belik S, Sareen J. Chronic pain conditions and suicidal ideation and suicide attempts: an epidemiologic perspective. Clin J Pain Mar-Apr 2008;24(3):204-210. [doi: 10.1097/ajp.0b013e31815ca2a3]
8. Braden JB, Sullivan MD. Suicidal thoughts and behavior among adults with self-reported pain conditions in the national comorbidity survey replication. J Pain 2008 Dec;9(12):1106-1115 [FREE Full text] [doi: 10.1016/j.jpain.2008.06.004] [Medline: 19038772]

9. Ilgen MA, Zivin K, McCammon RJ, Valenstein M. Pain and suicidal thoughts, plans and attempts in the United States. Gen Hosp Psychiatry 2008 Nov;30(6):521-527 [FREE Full text] [doi: 10.1016/j.genhosppsych.2008.09.003] [Medline: 19061678]
10. Campbell G, Darke S, Bruno R, Degenhardt L. The prevalence and correlates of chronic pain and suicidality in a nationally representative sample. Aust N Z J Psychiatry 2015 Sep 19;49(9):803-811. [doi: 10.1177/0004867415569795] [Medline: 25698809]
11. Smith M, Edwards R, Robinson R, Dworkin R. Suicidal ideation, plans, and attempts in chronic pain patients: factors associated with increased risk. Pain 2004 Sep;111(1-2):201-208. [doi: 10.1016/j.pain.2004.06.016] [Medline: 15327824]
12. Racine M, Choinière M, Nielson W. Predictors of suicidal ideation in chronic pain patients: an exploratory study. Clin J Pain May 2014;30(5):371-378. [doi: 10.1097/ajp.0b013e31829e9d4d]
13. Pinquart M, Silbereisen RK, Fröhlich C. Life goals and purpose in life in cancer patients. Support Care Cancer 2009 Mar 1;17(3):253-259. [doi: 10.1007/s00520-008-0450-0] [Medline: 18449572]
14. Dezutter J, Casalin S, Wachholtz A, Luyckx K, Hekking J, Vandewiele W. Meaning in life: an important factor for the psychological well-being of chronically ill patients? Rehabil Psychol 2013 Nov;58(4):334-241 [FREE Full text] [doi: 10.1037/a0034393] [Medline: 24295525]
15. Kowal J, Wilson K, Henderson P, McWilliams L. Change in suicidal ideation after interdisciplinary treatment of chronic pain. Clin J Pain Jun 2014;30(6):463-471. [doi: 10.1097/ajp.0000000000000003]
16. Steger M. Experiencing meaning in life: optimal functioning at the nexus of well-being, psychopathology, and spirituality. In: Wong P, editor. The Human Quest for Meaning: Theories, Research, and Applications, 2nd ed. New York: Routledge; 2012:165-184.
17. Steger M. Meaning in life. In: Lopez S, Snyder C, editors. The Oxford Handbook of Positive Psychology, 2nd Ed. Oxford: Oxford University Press; 2009:679-688.
18. Steger MF, Oishi S, Kesebir S. Is a life without meaning satisfying? The moderating role of the search for meaning in satisfaction with life judgments. J Positive Psychol 2011 Sep 26;6(3):173-180. [doi: 10.1080/17439760.2011.569171]
19. Costanza A, Prelati M, Pompili M. The meaning in life in suicidal patients: the presence and the search for constructs. A systematic review. Medicina (Kaunas) 2019 Aug 11;55(8):465 [FREE Full text] [doi: 10.3390/medicina55080465] [Medline: 31405240]
20. Steger M, Kashdan T. Search for meaning in life. In: Baumeister R, Vohs K, editors. Encyclopedia of Social Psychology. Thousand Oaks: Sage Publications; 2007:783-785.
21. Frankl V, Crumbaugh J. Psychotherapy and Existentialism: Selected Papers on Logotherapy, 4th Ed. New York: Simon and Schuster; 1967.
22. Baumeister R. Meanings of Life. New York: Guilford Press; 1991.
23. Steger MF, Frazier P, Oishi S, Kaler M. The meaning in life questionnaire: assessing the presence of and search for meaning in life. J Couns Psychol 2006 Jan;53(1):80-93. [doi: 10.1037/0022-0167.53.1.80]
24. Steger MF, Kashdan TB, Sullivan BA, Lorentz D. Understanding the search for meaning in life: personality, cognitive style, and the dynamic between seeking and experiencing meaning. J Pers 2008 Apr;76(2):199-228. [doi: 10.1111/j.1467-6494.2007.00484.x] [Medline: 18331281]
25. Dezutter J, Luyckx K, Wachholtz A. Meaning in life in chronic pain patients over time: associations with pain experience and psychological well-being. J Behav Med 2015 Apr 24;38(2):384-396 [FREE Full text] [doi: 10.1007/s10865-014-9614-1] [Medline: 25537924]
26. Lapierre S, Dubé M, Bouffard L, Alain M. Addressing suicidal ideations through the realization of meaningful personal goals. Crisis 2007 Jan;28(1):16-25. [doi: 10.1027/0227-5910.28.1.16] [Medline: 17555029]
27. Denneson LM, Teo AR, Ganzini L, Helmer DA, Bair MJ, Dobscha SK. Military veterans' experiences with suicidal ideation: implications for intervention and prevention. Suicide Life Threat Behav 2015 Aug 03;45(4):399-414. [doi: 10.1111/sltb.12136] [Medline: 25367753]
28. Kim HM, Levine DS, Pfeiffer PN, Blow AJ, Marchiondo C, Walters H, et al. Postdeployment suicide risk increases over a 6-month period: predictors of increased risk among midwestern Army National Guard soldiers. Suicide Life Threat Behav 2017 Aug 05;47(4):421-435. [doi: 10.1111/sltb.12303] [Medline: 27704587]
29. Marco JH, Pérez S, García-Alandete J. Meaning in life buffers the association between risk factors for suicide and hopelessness in participants with mental disorders. J Clin Psychol 2016 Jul 14;72(7):689-700. [doi: 10.1002/jclp.22285] [Medline: 26990856]
30. Costanza A, Chytas V, Mazzola V, Piguet V, Desmeules J, Bondolfi G, et al. The Role of Demoralization and Meaning in Life (DEMIL) in influencing suicidal ideation among patients affected by chronic pain: protocol of a single-center, observational, case-control study. JMIR Res Protoc 2020 Nov 26;9(11):e24882 [FREE Full text] [doi: 10.2196/24882] [Medline: 33144275]
31. Beck A, Steer R, Brown G. Beck Depression Inventory (BDI-II). London: Pearson; 1996.
32. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59 Suppl 20:22-33. [Medline: 9881538]

33. Beck AT, Kovacs M, Weissman A. Assessment of suicidal intention: the Scale for Suicide Ideation. J Consult Clin Psychol 1979;47(2):343-352. [doi: 10.1037/0022-006x.47.2.343]
34. Baertschi M, Costanza A, Canuto A, Weber K. The dimensionality of suicidal ideation and its clinical implications. Int J Methods Psychiatr Res 2019 Mar 13;28(1):e1755 [FREE Full text] [doi: 10.1002/mpr.1755] [Medline: 30426604]
35. Sokero TP, Melartin TK, Rytsälä H, Leskelä U, Lestelä-Mielonen P, Isometsä E. Suicidal ideation and attempts among psychiatric patients with major depressive disorder. J Clin Psychiatry 2003 Sep 15;64(9):1094-1100. [doi: 10.4088/jcp.v64n0916] [Medline: 14628986]
36. de Beurs DP, Fokkema M, O'Connor R. Optimizing the assessment of suicidal behavior: the application of curtailment techniques. J Affect Disord 2016 May 15;196:218-224. [doi: 10.1016/j.jad.2016.02.033] [Medline: 26938964]
37. Steger MF. Meaning and Purpose. Laboratory for the Study of Meaning and Quality of Life. 2019 Aug 25. URL: http://www.michaelfsteger.com
38. Harris C, D'Eon J. Psychometric properties of the Beck Depression Inventory-second edition (BDI-II) in individuals with chronic pain. Pain 2008 Jul 31;137(3):609-622. [doi: 10.1016/j.pain.2007.10.022] [Medline: 18079063]
39. Corbière M, Bonneville-Roussy A, Franche R, Coutu M, Choinière M, Durand M, et al. Further validation of the BDI-II among people with chronic pain originating from musculoskeletal disorders. Clin J Pain 2011;27(1):62-69. [doi: 10.1097/ajp.0b013e3181eef903]
40. Bourque P, Beaudette D. Étude psychometrique du questionnaire de dépression de Beck auprès d'un échantillon d'étudiants universitaires francophones. Can J Behav Sci 1982;14(3):211-218. [doi: 10.1037/h0081254]
41. Bouvard M, Cottraux J. Protocoles et échelles d'évaluation en psychiatrie et psychologie. Paris: Masson; 2010.
42. Wesley AL, Gatchel RJ, Garofalo JP, Polatin PB. Toward more accurate use of the Beck Depression Inventory with chronic back pain patients. Clin J Pain 1999 Jun;15(2):117-121. [doi: 10.1097/00002508-199906000-00008] [Medline: 10382925]
43. Choi Y, Mayer TG, Williams MJ, Gatchel RJ. What is the best screening test for depression in chronic spinal pain patients? Spine J 2014 Jul 01;14(7):1175-1182. [doi: 10.1016/j.spinee.2013.10.037] [Medline: 24225008]
44. Duburcq A, Blin P, Charpak Y, Blachier C, Allicar MP, Bouhassira M, et al. [Use of a structured diagnostic interview to identify depressive episodes in an epidemiologic study: a posteriori internal validation]. Rev Epidemiol Sante Publique 1999 Oct;47(5):455-463. [Medline: 10587996]
45. de Man AF, Balkou S, Iglesias R. Une version canadienne française du sondage sur les expériences vécues. Santé mentale au Québec 1987;12(2):181. [doi: 10.7202/030412ar]
46. de Man AF, Leduc CP, Labrèche-Gauthier L. A French-Canadian scale for suicide ideation for use with adolescents. Can J Behav Sci 1993;25(1):126-134. [doi: 10.1037/h0078786]
47. Costanza A, Baertschi M, Richard-Lepouriel H, Weber K, Pompili M, Canuto A. The presence and the search constructs of meaning in life in suicidal patients attending a psychiatric emergency department. Front Psychiatry 2020 Apr 28;11:327 [FREE Full text] [doi: 10.3389/fpsyt.2020.00327] [Medline: 32425828]
48. Costanza A, Chytas V, Piguet V, Luthy C, Mazzola V, Bondolfi G, et al. Meaning in life among patients with chronic pain and suicidal ideation: mixed methods study. JMIR Form Res 2021 Jun 04;5(6):e29365 [FREE Full text] [doi: 10.2196/29365] [Medline: 34003136]
49. Dewitte L, Vandenbulcke M, Dezutter J. Meaning in life matters for older adults with Alzheimer's disease in residential care: associations with life satisfaction and depressive symptoms. Int Psychogeriatr 2019 Feb 06;31(5):607-615. [doi: 10.1017/s1041610218002338]
50. Kleiman EM, Beaver JK. A meaningful life is worth living: meaning in life as a suicide resiliency factor. Psychiatry Res 2013 Dec 30;210(3):934-939. [doi: 10.1016/j.psychres.2013.08.002] [Medline: 23978733]
51. Lu H, Sheng W, Liao S, Chang N, Wu P, Yang Y, et al. The changes and the predictors of suicide ideation and suicide attempt among HIV-positive patients at 6-12 months post diagnosis: a longitudinal study. J Adv Nurs 2019 Mar 21;75(3):573-584. [doi: 10.1111/jan.13883] [Medline: 30334591]
52. Kleiman EM, Adams LM, Kashdan TB, Riskind JH. Gratitude and grit indirectly reduce risk of suicidal ideations by enhancing meaning in life: evidence for a mediated moderation model. J Res Pers 2013 Oct;47(5):539-546. [doi: 10.1016/j.jrp.2013.04.007]
53. McCracken LM, Morley S. The psychological flexibility model: a basis for integration and progress in psychological approaches to chronic pain management. J Pain 2014 Mar;15(3):221-234. [doi: 10.1016/j.jpain.2013.10.014] [Medline: 24581630]
54. Tighe J, Nicholas J, Shand F, Christensen H. Efficacy of acceptance and commitment therapy in reducing suicidal ideation and deliberate self-harm: systematic review. JMIR Ment Health 2018 Jun 25;5(2):e10732 [FREE Full text] [doi: 10.2196/10732] [Medline: 29941419]
55. Downs MF, Eisenberg D. Help seeking and treatment use among suicidal college students. J Am Coll Health 2012 Feb;60(2):104-114. [doi: 10.1080/07448481.2011.619611] [Medline: 22316407]
56. Michel K, Valach L, Gysin-Maillart A. A novel therapy for people who attempt suicide and why we need new models of suicide. Int J Environ Res Public Health 2017 Mar 01;14(3):243 [FREE Full text] [doi: 10.3390/ijerph14030243] [Medline: 28257071]

57. Schnyder U, Valach L, Bichsel K, Michel K. Attempted suicide. Gen Hosp Psychiatr 1999 Jan;21(1):62-69. [doi: 10.1016/s0163-8343(98)00064-4]
58. Racine M. Chronic pain and suicide risk: a comprehensive review. Prog Neuropsychopharmacol Biol Psychiatry 2018 Dec 20;87(Pt B):269-280 [FREE Full text] [doi: 10.1016/j.pnpbp.2017.08.020] [Medline: 28847525]
59. Webster J. Self-report wisdom measures: strengths, limitations, and future directions. In: Sternberg R, Glück J, editors. The Cambridge Handbook of Wisdom. Cambridge: Cambridge University Press; 2019:297-320.
60. Lorenzi-Cioldi F. Questions de méthodologie en sciences sociales. Lausanne et Paris: Delachaux & Niestlé; 1997.
61. Lau F. Chapter 12: Methods for correlational studies. In: Lau F, Kuziemsky C, editors. Handbook of eHealth Evaluation: An Evidence-based Approach. Victoria: University of Victoria; 2017.
62. Costanza A, Baertschi M, Richard-Lepouriel H, Weber K, Berardelli I, Pompili M, et al. Demoralization and its relationship with depression and hopelessness in suicidal patients attending an emergency department. Int J Environ Res Public Health 2020 Mar 26;17(7):2232 [FREE Full text] [doi: 10.3390/ijerph17072232] [Medline: 32225017]
63. Gysin-Maillart A, Schwab S, Soravia L, Megert M, Michel K. A novel brief therapy for patients who attempt suicide: a 24-months follow-up randomized controlled study of the Attempted Suicide Short Intervention Program (ASSIP). PLoS Med 2016 Mar;13(3):e1001968 [FREE Full text] [doi: 10.1371/journal.pmed.1001968] [Medline: 26930055]
64. Park A, Gysin-Maillart A, Müller TJ, Exadaktylos A, Michel K. Cost-effectiveness of a brief structured intervention program aimed at preventing repeat suicide attempts among those who previously attempted suicide: a secondary analysis of the ASSIP randomized clinical trial. JAMA Netw Open 2018 Oct 05;1(6):e183680 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.3680] [Medline: 30646253]
65. Gysin-Maillart A, Soravia L, Schwab S. Attempted suicide short intervention program influences coping among patients with a history of attempted suicide. J Affect Disord 2020 Mar 01;264:393-399. [doi: 10.1016/j.jad.2019.11.059] [Medline: 31759660]
66. Ring M, Gysin-Maillart A. Patients' satisfaction with the therapeutic relationship and therapeutic outcome is related to suicidal ideation in the Attempted Suicide Short Intervention Program (ASSIP). Crisis 2020 Sep;41(5):337-343. [doi: 10.1027/0227-5910/a000644] [Medline: 31918584]

Abbreviations

ACT: acceptance and commitment therapy
BDI-II: Beck Depression Inventory, 2nd Edition
DSM-IV: Diagnostic and Statistical Manual of Mental Disorders, 4th Edition
MiL: meaning in life
MiLQ: Meaning in Life Questionnaire
MINI: Mini-International Neuropsychiatric Interview
MPC: Multidisciplinary Pain Center
SA: suicidal attempt
SB: suicidal behavior
SI: suicidal ideation
SSI: Scale for Suicide Ideation
VAS: visual analog scale

Edited by A Mavragani; submitted 24.11.21; peer-reviewed by B Thompson, M Martelli, A Joseph; comments to author 11.01.22; revised version received 06.03.22; accepted 10.04.22; published 13.06.22.

Please cite as:
Chytas V, Costanza A, Mazzola V, Luthy C, Galani V, Bondolfi G, Cedraschi C
Possible Contribution of Meaning in Life in Patients With Chronic Pain and Suicidal Ideation: Observational Study
JMIR Form Res 2022;6(6):e35194
URL: https://formative.jmir.org/2022/6/e35194
doi:10.2196/35194
PMID:35699974

Original Paper

Data Privacy Concerns Using mHealth Apps and Smart Speakers: Comparative Interview Study Among Mature Adults

Tanja Schroeder[1,2], BA, MA; Maximilian Haug[2], BEng, MSc; Heiko Gewald[2], BSc, MBA, EMBSc, PhD

[1]Centre for Health Systems and Safety Research, Australian Institute of Health Innovation, Macquarie University, Sydney, Australia

[2]Center for Research on Service Sciences, Neu-Ulm University of Applied Sciences, Neu-Ulm, Germany

Corresponding Author:
Tanja Schroeder, BA, MA
Centre for Health Systems and Safety Research
Australian Institute of Health Innovation
Macquarie University
75 Talavera Road
Sydney, NSW 2109
Australia
Phone: 61 2 9850 ext 6281
Email: tanja.schroeder@hdr.mq.edu.au

Abstract

Background: New technologies such as mobile health (mHealth) apps and smart speakers make intensive use of sensitive personal data. Users are typically aware of this and express concerns about their data privacy. However, many people use these technologies although they think their data are not well protected. This raises specific concerns for sensitive health data.

Objective: This study aimed to contribute to a better understanding of data privacy concerns of mature adults using new technologies and provide insights into their data privacy expectations and associated risks and the corresponding actions of users in 2 different data contexts: mHealth apps and smart speakers.

Methods: This exploratory research adopted a qualitative approach, engaging with 20 mature adults (aged >45 years). In a 6-month test period, 10 (50%) participants used a smart speaker and 10 (50%) participants used an mHealth app. In interviews conducted before and after the test period, we assessed the influence of data privacy concerns on technology acceptance, use behavior, and continued use intention.

Results: Our results show that although participants are generally aware of the need to protect their data privacy, they accept the risk of misuse of their private data when using the technology. Surprisingly, the most frequently stated risk was not the misuse of personal health data but the fear of receiving more personalized advertisements. Similarly, surprisingly, our results indicate that participants value recorded verbal data higher than personal health data.

Conclusions: Older adults are initially concerned about risks to their data privacy associated with using data-intensive technologies, but those concerns diminish fairly quickly, culminating in resignation. We find that participants do not differentiate between risky behaviors, depending on the type of private data used by different technologies.

(JMIR Form Res 2022;6(6):e28025) doi:10.2196/28025

KEYWORDS

data privacy concerns; privacy paradox; mHealth app; smart speaker; mature adults; smartphone

Introduction

Overview

A mobile health (mHealth) app is a specific type of digital health app that uses mobile devices such as smartphones and tablets that are already integrated into daily lives of people. People use mHealth apps to monitor their health or access medical information or assistance through wireless mobile devices such as smartphones and portable monitoring devices [1]. Similarly, mHealth apps enable health care providers to monitor certain user activities and behaviors so that they can provide personalized health care advice. Other technologies can also be used to support personalized health care support. For example, smart speakers with artificial intelligence–operated assistants that help users more easily access information or control other devices via their voice can be used for this purpose [2,3].

Although smart speakers are often used for everyday activities, such as playing music, setting a timer, or hearing a weather report, they can also be used to remind users to take their medication or answer health-related questions. Smart speakers collect and process a variety of private information and are typically used in private environments. Because a smart speaker must be able to recognize the voice activation keyword at any time, its microphone's default status is active. Therefore, many people associate smart speakers with involuntary personal information disclosure [4].

The increasing use of digital and mobile technologies, combined with the need for personalized and cost-efficient health care, has fostered the emergence of mHealth technologies. Such technologies have many potential health care benefits, such as the ability to monitor users' health status continuously and remotely, increased diagnostic accuracy, earlier awareness of new problems, lower health care costs, greater availability of health care to people living in remote areas, and improved doctor-patient communications.

Despite these potential benefits, the sensitivity of individuals' private data collected and recorded by these digital apps raises concerns about the privacy of this information [5,6]. For example, some potential users want to control what people in their private environment, such as family members, know about their health status, perhaps because they fear being judged, reprimanded, discriminated against, or even penalized for their physical and health status [7]. Some people may not want to take care of their family members because of their current health status [7,8]. The fear of social stigmatism is another reason people may not want others to know their health status [8]. This may apply to physical or mental disabilities, mental illnesses, or certain diseases such as HIV and Alzheimer [8,9].

In this study, we define *privacy* as the right of individuals, groups, or institutions to determine when, how, and to what extent information about them is shared with others [10]. Privacy is a subjective concept linked to an individual's perception of what constitutes a threat to their personal property or physical or moral integrity, depending on cultural aspects and sociodemographic issues [11]. Users' perspectives on interactions and communications influence their data privacy–related decisions on a range of privacy issues, including technical issues such as regulating visibility in social networks and using smartphone apps that collect confidential data [12].

Most extant data privacy literature focusing on mHealth deals with the technical aspects of privacy, such as the level of security of information transmitted over mobile networks and stored on a device or in a cloud service needed to prevent unauthorized access to a patient's information [8,13-15]. However, privacy is not only a technical issue. For example, the collaborative use of mHealth apps for shared care management imposes other privacy requirements related to human factors, such as the wishes and preferences of the user when exchanging health information with authorized institutions and external persons such as professional health care providers [16].

Extant research shows that privacy concerns are a major inhibitor of the adoption and use of both mHealth apps and smart speakers [4,17]. Because the 2 technologies access different types of sensitive personal data, potential users may have different privacy concerns regarding each technology. To investigate this issue, we posed the following research questions (RQs):

1. RQ1: What privacy concerns do potential users associate with mHealth apps and smart speakers?
2. RQ2: What data privacy–related risks do potential users attribute to the use of mHealth apps and smart speakers?
3. RQ3: What privacy-related issues lead to rejection of mHealth apps and smart speakers?

Although the use of smart speakers is roughly equal among *mature adults*, which we define as people who are aged >45 years and adults aged <45 years [18], extant research shows that mobile apps for health care purposes are most commonly used by mature adults [19]. Because this study addresses both technologies, we focus on the mature adult user group, providing 50% (10/20) of the participants with a smart speaker and 50% (10/20) of the participants with an mHealth app to use in a 6-month testing phase. All the participants were interviewed before and after the testing phase.

Background

One of the main reasons why people are reluctant to use mHealth apps is concerns about the security and privacy of their health-related data [20-23]. Users often do not know what kind of data mHealth apps collect and store and who can access data entered manually or collected by sensors and for what purposes [20,24]. Studies show that users have greater security and privacy concerns about mHealth apps that focus on issues associated with stigmatization, discrimination, or social isolation, such as sexually transmitted diseases, sexual orientation, and mental illnesses [25-29]. Considering that millions of patients' health data have been compromised through hacking or other incidents in recent years, these concerns are valid [30]. Despite private data security breaches, few mHealth apps have security features that adequately protect users' private health data [31-33].

Privacy Theories

Extant research has not yet fully explored the specific role of privacy concerns in the acceptance and use of technology. The privacy calculus theory [34,35] and the so-called *privacy paradox* [36] are central concepts in privacy research. They illustrate the ambivalent influence of privacy on behavior.

The privacy calculus theory assumes that individuals engage cognitively to weigh the perceived costs and benefits of a behavior [34,35]. If the benefits outweigh the potential harm that privacy abuse can induce, individuals engage with the technology. However, this view is partly challenged by the privacy paradox, a phenomenon in which individuals engage with a technology even though the privacy concerns they associate with using the technology outweigh the anticipated benefits of using it [36,37]. Although both phenomena are well documented in practice, technology adoption scholars have yet to explain this conundrum [38].

Privacy Research

Computer system privacy has long been a concern. In 1969, a study by Hoffman [39] discussed strategies for user access control and data protection, emphasizing the need to weigh the efficiency benefits of storing personal information against the risks of third-party access to such information. The study examined legal and administrative safeguards to protect sensitive information on computers and evaluated the technical solutions available at the time.

More recently, a study by Barth and De Jong [40] focused on privacy-related human factors. They conducted a systematic literature review to understand the web-based privacy paradox, where users indicate great concern about the privacy of their personal information but do very little to protect such data. The authors identified 35 theoretical approaches to decision-making and identified different perspectives on the paradox. Specifically, they discuss the decision-making process after rational or irrational risk-benefit calculations in the specific context of the privacy paradox.

Pavlou [41] described the data protection paradox in his privacy briefings as a phenomenon in which individuals express strong concerns about their privacy but behave in a way that contradicts these concerns. For example, some consumers still share their personal information despite privacy concerns. Even after considering the user perspective, Aïmeur [42] addressed the question of how to achieve a good compromise between privacy and user personalization. He mentioned that an increasing number of users can only control their data by fine-tuning the app settings. The author also argued that mHealth would benefit significantly if users had direct control over when, where, and with whom their personal data were shared.

Privacy-Personalization Paradox

Varshney [43] described privacy as the right of a group or individual to isolate or retain information about themselves. Personalization technologies offer users a wide range of services from which to choose but also require users to disclose more personal information, which may raise privacy concerns [44]. This has been compounded by the emergence of smartphones that can capture personal information more accurately [45]. For example, health care counseling provided via mobile platforms can reduce the need for personal interaction, but app users must then share information relevant to their health, such as health status, preferences, and lifestyle, as well as their telephone number with service providers to use personalized health care counseling services that overcome geographic barriers and save time. This, in turn, raises privacy concerns regarding the collection of sensitive consumer information: a technological paradox [46]. Although consumers want personalized services, they are reluctant to disclose personal information and want to disclose as little information as possible.

Demographics

Demographic differences between potential consumers are linked to behavioral intentions [47-50]. Some studies have focused on age differences in technology adoption, suggesting that there are differences in intentional behavior among different age groups [51]. However, Featherman and Pavlou [52] found that the validity of theoretical constructs, including models of health behavior change, is not well documented at all life stages. Most scholars agree that as people age, their physical and mental activities change, which affects their health status and decision-making [53]. Researchers have recently recognized that studying age differences in behavioral intention in the health context is both useful and essential. Ziefle and Röcker [54] found that age differences played an important role in the acceptance of health-related technologies. Similarly, Sintonen and Immonen [55] found that older participants' intention to adopt technology varies over time and according to the service provided, whereas a study by Guo et al [56] found that *dark-side* constructs influence older participants' intention to adopt mHealth services.

Methods

Recruitment and Demographics

This is an exploratory qualitative study with 20 participants divided into 2 groups. Semistructured interviews were conducted between October 2019 and April 2020, in southern Germany. Each group consisted of 10 participants aged between 46 and 80 years. Group A participants used smart speakers with Amazon Alexa technology, and group B participants used a self-provided mHealth app on their Android smartphone. Both technologies were provided to the participants free of charge.

Smart speakers are voice-controlled, enabling users to play music or news, access information, place telephone calls, and perform other tasks via voice commands. The smart speaker is activated by speaking a predefined voice command. The smart speakers have several levels of built-in privacy measures. For example, a microphone can be deactivated by pushing a button that interrupts its power supply. Users also retain full control over voice recordings. Users can also control whether, when, and which voice recordings are accessed or accessible by third parties through the internet.

The mHealth app incorporates a health diary that users can use to track their daily health status, water intake, and other health-related information. This app is mainly intended to make daily life easier for older people. Referring to a digital health diary should allow health care providers to collect relevant information more effectively and quickly and make the process less burdensome for older people. In addition, it aims to prevent a decline in cognitive performance. Using the *My values* feature, users can choose from a wide range of functions within the app, including body temperature, blood pressure, heart rate, blood glucose levels, and weight. Users can query the recorded data using various time intervals. In addition, users can activate a push notification function to receive reminders to enter current measurement values.

The participants were recruited in southern Germany through flyers and posters in senior citizen centers and postings on social media. Participation was strictly voluntary, and no incentives were provided. All participants interviewed were informed of the research team's data protection arrangements and signed a document to this effect in full compliance with the European Union data protection regulations, the General Data Protection

Regulation (GDPR). Participants interested in taking part in the study were given a participant information sheet and an informed consent form detailing the participation requirements in advance, with the option to have the form explained if necessary. Participants signed and returned the form, indicating their informed consent. The consent forms were retained and stored securely as a record of informed consent. Participants could withdraw from the interviews at any time during the study if they no longer wanted to be included. Furthermore, the researchers were available to answer questions at any time. The research group saved all participant data anonymously and retained all written and audio materials collected during the interviews until the end of the data retention period. All project data are stored electronically on secure password-protected servers and accessible only by designated and approved project team members. All the data will be destroyed at the end of the data retention period.

Table 1 provides demographic information of the participants.

All participants were considered IT affine, demonstrating interest in digital technologies such as smartphones, tablets, PCs, smart watches, or smart speakers. All participants actively used a smartphone and at least one social media service regularly and demonstrated at least an average internet use intensity for their age group. All interviewees were able to install and set up a smart speaker or an mHealth app with little or no assistance.

Table 1. Demographics of the participants (N=20).

Participant group and informant	Age (years)	Gender
Group A		
A1	56	Female
A2	59	Female
A3	54	Female
A4	48	Female
A5	49	Male
A6	70	Male
A7	68	Male
A8	56	Male
A9	54	Male
A10	46	Male
Group B		
B1	52	Male
B2	50	Male
B3	55	Female
B4	60	Female
B5	56	Female
B6	61	Female
B7	54	Female
B8	52	Female
B9	80	Male
B10	51	Male

Data-Gathering Process

The study was structured into three phases: (1) before, (2) during, and (3) after testing the respective technologies.

Phase 1 began with participant recruitment and ended with the delivery of the technology to be tested. This phase included the first conversation to give participants detailed information about the study approach and timeline, collect demographic data, interview all participants about their expectations toward using new technology and their concerns about protecting their private data, and ask participants who would test the mHealth app about their experiences in preventive health care. We also asked specific questions about privacy concerns related to the technology tested in our study.

Phase 2 started with the delivery of the technology and ended when the technology was returned (smart speaker) or uninstalled (mHealth app). This 6-month phase gave participants ample time to test the technology thoroughly. During this phase, research assistants (MH, Jennifer Klaus, and Jaro Lanza) were available to help if participants encountered difficulties using the technology.

Phase 3 started when the technology was returned or uninstalled and ended when all data were collected and ready to be

analyzed. During this phase, we conducted posttest interviews with each participant to discuss their use behavior, observations, problems, and concerns. We paid special attention to how their assessment of data privacy concerns changed over time and how this influenced their intention to continue using the technology.

Interview Structure and Data Analysis Method

In phases 1 and 3, we conducted individual semistructured interviews with each participant, following an interview guide and drawing on a list of topics and specific questions. We posed open-ended questions that allowed the respondents to explore their experiences and views. The interview guide helped focus the interview and ensured the comparability of data collected among multiple participants in various interview settings and by different researchers. The interview process was systematic and comprehensive, but the interviewer was free to follow up on issues of greater interest or importance to the participant during the interview by adapting preformulated questions ad hoc to gain a deeper and more holistic understanding of the participants' perspectives and perceptions.

As the aim of this study was to understand the relationships among privacy concerns, risk perception, and use behavior in the technology-use context, we focused principally on privacy concerns, the impact of data misuse, and termination of use.

The 10 pretest and the 10 posttest interviews lasted 15 to 60 minutes each (average 35, SD 15.59 minutes) and were conducted face to face or via telephone or videoconference. The interviews were conducted in German and recorded, transcribed, and coded using an open coding approach using NVivo (version 10; QSR International) software independently by 2 research team members (TS and MH). The research team coded the data parallel to data collection. Subsequently, the data were triangulated according to the recommendations by Miles et al [57] and Flick [58]. The analysis took an inductive and interpretative approach. The inductive approach is a systematic procedure for analyzing qualitative data in which the analysis is guided by specific objectives [59].

In the second round of analysis, the codes from the individual interviews were correlated [57] to cast light on the specific characteristics of these topics and the influence of these factors in the context of the two technologies.

Ethics Approval

This study did not require ethical approval according to the guideline of the applicable Ethics Committee of the Bavarian Universities (Gemeinsame Ethikkommision der Hochschulen Bayerns [60]), as no risks or harm to the participants were expected and the basic ethical principles were not violated. All participants received a participant information and consent form explaining the requirements for participation, with the option to have the form explained to them if needed and gave their verbal consent as a sign of informed consent if they were willing to participate at the time of the interview. They were also given the opportunity to complete and sign the participant consent form to indicate their agreement to the interview being conducted.

Results

Overview

The pretest interviews focused on participants' general expectations regarding the technology, as well as their privacy concerns and expected risks associated with using the technology. The posttest interviews focused generally on how the informants had used the technology during the test phase, specifically on whether their privacy concerns and their perceptions of the risks changed during the 6-month testing phase, and if the intention to continue using the technology.

To better segregate the 2 different types of data used by the technologies we refer to "health data" for the data used by the mHealth app which predominantly consist of the participant's health status or well-being information and "personal data" for the data collected by the smart speaker, which comprises mainly intended and unintended speech, as well as the information transmitted when the user gives commands to the speaker.

Privacy Concerns

Overview

Individuals' privacy concerns are shaped and influenced by many factors, such as personal experiences, media coverage, and the social environment [61-64]. Therefore, it is essential to understand participants' perceptions of individual data privacy–related issues to be able to classify their statements and derive results. To better understand how privacy concerns impact the use of smart speakers and mHealth apps, we asked participants directly about their privacy concerns and perceptions of privacy-related issues.

It should be noted that some of the participants testing a smart speaker had previous experience using them, were more aware of privacy-related issues associated with them, and had already formed opinions about the risks and benefits of using them. In contrast, the participants testing a smart speaker for the first time had a basic understanding of the technology but were less aware of the privacy-related issues associated with them and did not know what to expect. None of the participants testing the mHealth app had prior experience of using an mHealth app.

The following 2 sections present our core findings, supported by exemplary quotations translated from the German original. The alphanumeric codes after each quotation refer to the quoted participants and other participants who expressed similar viewpoints.

Smart Speakers

Participants commonly expressed awareness of privacy issues associated with using a smart speaker:

> *I would say that I have a high level of data protection awareness. [...] I am very aware that the data that I enter on the Internet or that is processed via the Internet can be accessed by providers and misused.* [A8; A10]

Most (8/10, 80%) participants expressed concerns regarding data protection:

> *No, I do not think that the data is secure.* [A2; A5; A8; A9]
>
> *I do not think the data is secure. The data is recorded and stored, and once the data is on the Internet, it is not safe for me either.* [A3]

In addition, most (6/10, 60%) participants articulated some degree of concern that their data would be stored somewhere unknown and used without their permission. This illustrates a general tendency of users to doubt that their data will be protected:

> *I think there are loopholes and problems, and that data is not protected adequately.* [A1; A5; A6; A7; A8; A9]

Most (6/10, 60%) participants expressed resignations. Although they were concerned about their privacy, they also recognized that if they wish to participate in web-based activities or use certain apps, they must accept the terms of use and may lose control over their data. Many eventually decide to use an app, while maintaining some degree of privacy awareness:

> *Let me put it like this: I have a rather low data protection awareness. I ignore who can process my data.* [A1; A2]
>
> *I would describe my data protection awareness as not consistent enough to protect my data.* [A3; A5]
>
> *I take care of my data. Nevertheless, I think that I cannot really influence or intervene and determine who gets my data. As soon as I download and use an app, I have to agree to the terms of use.* [A6; A7]

However, some (2/10, 20%) participants expressed no concerns at all. A prominent issue is confidence in the manufacturer (in this case, Amazon).

> *The way I see it, as long as I feel sure, I will use the device; if trust is no longer there, then I think that I will no longer use the device.* [A6]
>
> *I have great confidence in Amazon that my data will be stored securely.* [A10]

In summary, our results show that most (8/10, 80%) participants have privacy concerns about using a smart speaker and are not sure what happens to their data. However, the participants did not fear monetary loss or reputational injury. We found that participants were willing to suppress their own data protection concerns to use the device and justified putting aside any lack of trust in the provider.

mHealth App

In the group using the mHealth app, participants' views on privacy and data protection concerns were split—half (5/10, 50%) of the participants said that they were not worried or concerned about the privacy of their personal data:

> *They cannot do much with my data anyway because I am an ordinary person. [...] what I use or look at or, how should I say it, [...] that is what every human being does, to put it like that.* [B1]
>
> *No. I've got nothing to hide.* [B5]
>
> *Not at all, actually. [...] And I have nothing to hide. But I am not afraid that this will end up anywhere.* [B9]

In contrast, (4/10, 40%) of the participants were concerned about data protection and privacy:

> *Yes, I have become very concerned. [...] Because the different sites are obviously not safe and a lot of data is collected about you, you do not know anything about it. Of course, it is great to have computers here, if you get a lot of information, but there is also a certain potential to become dependent on them, and as I said, I see a problem with surveillance and abuse.* [B6]

The most frequently mentioned concerns include surveillance, abuse, and use of your data against you:

> *Partly yes, but I mean I know what I can write about on my smartphone and what I cannot say. [...] I think you have to have restrictions about that. [...] I am just careful what I [...] write. [...] What would be on a postcard, you would write like that, I would say.* [B10]

Overall, our results show that participants vary in terms of their data protection concerns and how they engaged with data privacy, specifically with regard to using a smart speaker or an mHealth app.

Smart Speaker Versus mHealth App

Our results indicate that attitudes vary depending on the device and how information is captured or entered. As expected, participants were most sensitive to personal data. Somewhat surprisingly, our results indicate that participants are more concerned about personal speech recorded by a smart speaker than about personal health data entered into an mHealth app.

Data Misuse

To better understand the degree to which participants' concerns about data misuse are justified, we inquired about the perceived ramifications of misuse of data collected via the smart speaker or eHealth app.

Smart Speaker

Most (9/10, 90%) participants testing smart speakers associate smart speaker data misuse with personalized advertisements. Some have linked it to profiling, monetary loss, and data loss:

> *I think that my data will be used for advertisement. I mean that they sell my data to agencies to show me ads at the right time.* [A7; A2]
>
> *They could take money from my bank account. Alternatively, they will spam me with e-mails or ads.* [A4; A10]

mHealth App

Of the 7 participants who reported believing that their data were sold to unknown third parties, 3 (43%) assumed that their data were stored on unknown servers by companies such as Google. A few (3/10, 30%) participants had never considered this issue:

> *They trade our data, the data is sold, and that is well known. And yes, it is used to send advertisements to advertising companies.* [B8]
>
> *The data is just sitting around in some archive. Or someone can buy it. No idea. I have not really thought about it.* [B4]

Some (2/10, 20%) participants raised concerns about potential negative implications for existing insurance policies or apps for new insurance policies because of chronic or serious illnesses:

> *Yes, I do not see any problems there now, because my health is good. I can well imagine that others might have a problem because they think if everyone knows that I have trouble taking out an insurance policy or get offered worse conditions, but theoretically this is already the case.* [B3]
>
> *No, not really. I no longer have a problem with any health insurance companies at my age. I am privately insured so of course nothing will change. And I do not need to take out large insurance policies anymore, but if I were younger, I would be much more aware of the fact that the data might be misused.* [B6]

These statements indicate a much deeper understanding of the potential implications of data abuse in the health care context, including direct monetary damage. However, the risks described did not result in grave concerns or technology rejection among the respondents.

Termination of Use

No participant in either group stopped using the technology or intended to stop using the technology after the test because of data privacy concerns, and we observed no increase in data privacy concerns over time. Across the board, participants assumed that data were collected to enable personalized advertisements, which did not represent a salient enough risk to participants to motivate the termination of use.

Smart Speakers

Participants testing smart speakers reported several data privacy issues that would hypothetically motivate them to stop using it, including privacy violations, data leaks, and eavesdropping:

> *I would stop using it if I myself were affected, such that people could openly access my data.* [P5; P7; P9]
>
> *If I suddenly got advertising/promotional mails or phone calls, I would not use it anymore.* [P1; P3; P4; P8]

However, such hypothetical concerns and risks did not motivate participants to discontinue using technology. Rather, participants would have to experience an incident personally to trigger actual termination of use.

mHealth App

Participants testing the mHealth app reported that no data privacy issues would hypothetically motivate them to stop using the app:

> *No, not at all.* [B3]
>
> *Not significantly.* [B9]

Interestingly, none of the participants were seriously concerned about data privacy and the protection of their health data when using the mHealth app.

Discussion

Principal Findings

This study aimed to determine whether data privacy is perceived differently for different data-intensive technologies. By comparing 2 groups of mature adult users, we assess how the impact of privacy concerns, such as data protection concerns and risks, on technology use and discontinuation differs depending on the technology.

In both user groups, none of the participants stopped using the technology, despite privacy concerns. This is in line with the widely discussed privacy paradox [36], which states that individuals engage with technology even when they associate it with potential privacy loss issues. This phenomenon has been observed in social media, e-commerce, and mobile apps [45,65,66]. A likely driver of this paradoxical behavior is that the perceived immediate benefits of using a technology outweigh its potential, hypothetical, future risks [67]. Our results also indicate that this applies to health data privacy risks as well. In both user groups, participants expressed widespread resignation that choosing to use the technology comes at the cost of a greater risk of potential loss of data privacy.

Surprisingly, our results also show that people do not associate greater risk with personal health data collected by an mHealth app than with spoken words recorded by smart speakers. Because personal health data are highly sensitive and provide deep insights into individuals, one would expect users to be more concerned about protecting their privacy [68]. One possible explanation is that discrete data entered manually, consciously, and willingly, such as health status data in an mHealth app, are easier to control than impromptu utterances spoken all day long in a private setting. Regardless, our results call into question whether users of information technology, who require personal information distinguish between the data types collected by various devices and whether users fully understand the financial and personal ramifications of data abuse beyond personalized advertising.

In contrast with the mHealth app, which only passively receives data entries, a smart speaker may be perceived as invasively and nontransparently intruding into the private space. Indeed, several smart speak testers referred to the device's constant *listening* for the activation keyword as *spying* in the pretest interviews. However, no participants used this term in the posttest interviews after the 6-month test phase. This may indicate that they perceived greater risk to their privacy in the preadoption than in the postadoption phases, as they are commonly referred to in technology adoption research [69,70].

Our results provide initial indications of a potential decline in the perceived fear of loss of private health data with increasing age. Several older participants stated that no one would be interested in their health data; therefore, privacy breaches posed no threat to them. Further research is needed to discern consistent patterns of age-related differences.

Implications for Research

This exploratory research contributes to research on how data privacy concerns influence the intended and continuous use of data-intensive technologies, specifically mHealth apps and smart speakers. Investigations into the specific perceptions and resulting behaviors of mature adults in privacy research are scarce, and scholars still have only a rudimentary understanding of how older people's engagement with technology is influenced by privacy-related issues and concerns. This is especially important, as mHealth apps and voice-activated assistants gain increasing importance in providing health care–related services to mature and older adults [71].

Our findings have 4 major implications for research. As our findings are based on an exploratory qualitative research method, their implications do not aim to be representative of all mature adults but provide a basis for further research to investigate quantitatively.

First, our findings indicated that mature adults' self-reported privacy concerns did not directly influence their actual use behavior once they had adopted a data-intensive technology. This finding is consistent with the privacy paradox [36]. Participants frequently entered a state of resignation, acknowledging that choosing to use the technology requires them to accept the potential loss of data privacy.

Second, our results indicated that mature adults sometimes view their personal data indiscriminately. Even when participants were aware of the risks of data misuse, they articulated that the risks did not affect them personally or would not manifest themselves in their case. The risk that participants mentioned the most was receiving more personalized advertisements. In fact, participants valued protecting their general personal data more than protecting their personal health data, positing that they were not worthwhile targets of health data theft. This is concerning, especially considering recent health-related data hacks [72].

Third, our results pointed to age as a moderating factor in the perception of data privacy, risk assessment, and the subsequent application of privacy calculus. Several older participants articulated that their data privacy–related risks were low because of their age. Lee et al [73] showed that the tendency to discount future risks was prevalent among younger people. Our study indicated that this influence may be more substantial for older people who, according to our study, lack motivation to engage with how their personal data will be treated and are, therefore, more willing to disclose their personal data.

Finally, somewhat surprisingly, we found that data type did not significantly affect how participants perceive data privacy issues. We expected that participants would be more concerned about protecting the privacy of their personal health data than about protecting more general personal information. However, participants initially considered all personal data as equally worthy of protection, possibly insufficiently understanding the ramifications of data abuse and not reliably distinguishing between different types of data processed through different devices. Further research is needed on the role of data types in technology users' privacy calculus, as current privacy models do not distinguish among different types of data or consider individuals' perceptions of different types of data.

Limitations and Further Research

This study is exploratory in nature; therefore, our results are not generalizable to all mature adults and all data-intensive technologies. Furthermore, all participants were recruited in the south of Germany and thus shared a certain cultural background. As with every qualitative study, this study is subject to potential bias from the research team and potentially influenced by social desirability bias among participants.

Our findings suggest several avenues for further research. Specifically, we call for further research on how the act of *resignation* manifests in users' privacy calculus as an acceptable price to pay for using a certain technology. Further research is also needed to understand what drives people to value certain types of personal data more than others, which, in our case, is valuing general personal data used to personalize advertising more than personal health data, which can be misused with grave consequences. Finally, a cross-generational study is required to assess the influence of age on data privacy concerns and technology adoption.

Conclusions

Research on how mature adults' data privacy concerns influence their use of data-intensive technology is scarce, despite reports of data hacks and leaks and eavesdropping on prominent technologies and the sensitive nature of personal health data. The results of our exploratory research analyzing interviews with 20 mature adult users of data-intensive technologies reveals that although participants self-reported initial data privacy concerns, they did not value the risks high enough to discontinue using the data-intensive technologies in focus. Rather, they expressed widespread resignation that choosing to use the technology means accepting the risk of loss of data privacy. This fatalistic surrender, combined with evidence that participants valued their general personal data more than their personal health data, is a cause for concern about the security of personal data among technology users of this generation.

Conflicts of Interest

None declared.

References

1. Global diffusion of eHealth: making universal health coverage achievable: report of the third global survey on eHealth. World Health Organization. 2017. URL: https://apps.who.int/iris/handle/10665/252529 [accessed 2012-11-20]

2. Park K, Park Y, Lee J, Ahn JH, Kim D. Alexa, Tell Me More! The effectiveness of advertisements through smart speakers. Int J Electron Commer 2022 Feb 16;26(1):3-24. [doi: 10.1080/10864415.2021.2010003]
3. Kim S. Exploring how older adults use a smart speaker-based voice assistant in their first interactions: qualitative study. JMIR Mhealth Uhealth 2021 Jan 13;9(1):e20427 [FREE Full text] [doi: 10.2196/20427] [Medline: 33439130]
4. Malkin N, Deatrick J, Tong A, Wijesekera P, Egelman S, Wagner D. Privacy attitudes of smart speaker users. Proc Priv Enh Technol 2019;2019(4):250-271. [doi: 10.2478/popets-2019-0068]
5. Boulos MN, Brewer AC, Karimkhani C, Buller DB, Dellavalle RP. Mobile medical and health apps: state of the art, concerns, regulatory control and certification. Online J Public Health Inform 2014 Feb 5;5(3):229 [FREE Full text] [doi: 10.5210/ojphi.v5i3.4814] [Medline: 24683442]
6. Dehling T, Gao F, Schneider S, Sunyaev A. Exploring the far side of mobile health: information security and privacy of mobile health apps on iOS and Android. JMIR Mhealth Uhealth 2015 Jan 19;3(1):e8 [FREE Full text] [doi: 10.2196/mhealth.3672] [Medline: 25599627]
7. Prasad A, Sorber J, Stablein T, Anthony D, Kotz D. Understanding sharing preferences and behavior for mHealth devices. In: Proceedings of the 2012 ACM Workshop on Privacy in the Electronic Society. 2012 Presented at: WPES '12; October 15, 2012; Raleigh, NC, USA p. 117-128. [doi: 10.1145/2381966.2381983]
8. Silva W, Sacramento C, Silva E, Garcia AC, Leal Ferreira SB. Health information, human factors and privacy issues in mobile health applications. In: Proceedings of the 53rd Hawaii International Conference on System Sciences. 2020 Presented at: HICSS '20; January 7-10, 2020; Maui, HI, USA p. 3429-3438. [doi: 10.24251/hicss.2020.420]
9. Koffi B, Yazdanmehr A, Mahapatra R. Mobile Health Privacy Concerns - A Systematic Review. In: Proceedings of the 24th Americas' Conference on Information Systems. 2018 Presented at: AMCIS '18; August 16-18, 2018; New Orleans, LA, USA p. 25 URL: https://aisel.aisnet.org/amcis2018/Health/Presentations/25
10. Westin AF. Privacy and freedom. Wash Lee L Rev 1968 Mar 1;25(1):165-170.
11. Dupuis MJ, Crossler RE, Endicott-Popovsky B. Measuring the human factor in information security and privacy. In: Proceedings of the 49th Hawaii International Conference on System Sciences. 2016 Presented at: HICSS '16; January 5-8, 2016; Koloa, HI, USA p. 3676-3685. [doi: 10.1109/hicss.2016.459]
12. Acquisti A, Adjerid I, Balebako R, Brandimarte L, Cranor LF, Komanduri S, et al. Nudges for privacy and security: understanding and assisting users' choices online. ACM Comput Surv 2018 May 31;50(3):1-41. [doi: 10.1145/3054926]
13. Sharma S, Chen K, Sheth AP. Toward practical privacy-preserving analytics for IoT and cloud-based healthcare systems. IEEE Internet Comput 2018 Mar;22(2):42-51. [doi: 10.1109/mic.2018.112102519]
14. Saleheen N, Chakraborty S, Ali N, Mahbubur Rahman M, Hossain SM, Bari R, et al. mSieve: differential behavioral privacy in time series of mobile sensor data. Proc ACM Int Conf Ubiquitous Comput 2016 Sep;2016:706-717 [FREE Full text] [doi: 10.1145/2971648.2971753] [Medline: 28058408]
15. Silva BM, Rodrigues JJ, Canelo F, Lopes IC, Zhou L. A data encryption solution for mobile health apps in cooperation environments. J Med Internet Res 2013 Apr 25;15(4):e66 [FREE Full text] [doi: 10.2196/jmir.2498] [Medline: 23624056]
16. Avancha S, Baxi A, Kotz D. Privacy in mobile technology for personal healthcare. ACM Comput Surv 2012 Nov;45(1):1-54. [doi: 10.1145/2379776.2379779]
17. Zhou L, Bao J, Watzlaf V, Parmanto B. Barriers to and facilitators of the use of mobile health apps from a security perspective: mixed-methods study. JMIR Mhealth Uhealth 2019 Apr 16;7(4):e11223 [FREE Full text] [doi: 10.2196/11223] [Medline: 30990458]
18. Kinsella B. Voice Assistant Demographic Data – Young Consumers More Likely to Own Smart Speakers While Over 60 Bias Toward Alexa and Siri. Voice Bot AI. 2019 Jun 21. URL: https://voicebot.ai/2019/06/21/voice-assistant-demographic-data-young-consumers-more-likely-to-own-smart-speakers-while-over-60-bias-toward-alexa-and-siri/ [accessed 2021-05-05]
19. Rasche P, Wille M, Bröhl C, Theis S, Schäfer K, Knobe M, et al. Prevalence of health app use among older adults in Germany: national survey. JMIR Mhealth Uhealth 2018 Jan 23;6(1):e26 [FREE Full text] [doi: 10.2196/mhealth.8619] [Medline: 29362211]
20. Atienza AA, Zarcadoolas C, Vaughon W, Hughes P, Patel V, Chou WY, et al. Consumer attitudes and perceptions on mHealth privacy and security: findings from a mixed-methods study. J Health Commun 2015;20(6):673-679. [doi: 10.1080/10810730.2015.1018560] [Medline: 25868685]
21. Kao CK, Liebovitz DM. Consumer mobile health apps: current state, barriers, and future directions. PM R 2017 May;9(5S):S106-S115. [doi: 10.1016/j.pmrj.2017.02.018] [Medline: 28527495]
22. Kotz D, Gunter CA, Kumar S, Weiner JP. Privacy and security in mobile health: a research agenda. Computer (Long Beach Calif) 2016 Jun;49(6):22-30 [FREE Full text] [doi: 10.1109/MC.2016.185] [Medline: 28344359]
23. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. JMIR Mhealth Uhealth 2015 Nov 04;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
24. Wiesner M, Zowalla R, Suleder J, Westers M, Pobiruchin M. Technology adoption, motivational aspects, and privacy concerns of wearables in the German running community: field study. JMIR Mhealth Uhealth 2018 Dec 14;6(12):e201 [FREE Full text] [doi: 10.2196/mhealth.9623] [Medline: 30552085]

25. Di Matteo D, Fine A, Fotinos K, Rose J, Katzman M. Patient willingness to consent to mobile phone data collection for mental health apps: structured questionnaire. JMIR Ment Health 2018 Aug 29;5(3):e56 [FREE Full text] [doi: 10.2196/mental.9539] [Medline: 30158102]
26. Goldenberg T, McDougal SJ, Sullivan PS, Stekler JD, Stephenson R. Preferences for a mobile HIV prevention app for men who have sex with men. JMIR Mhealth Uhealth 2014 Oct 29;2(4):e47 [FREE Full text] [doi: 10.2196/mhealth.3745] [Medline: 25355249]
27. Goldenberg T, McDougal SJ, Sullivan PS, Stekler JD, Stephenson R. Building a mobile HIV prevention app for men who have sex with men: an iterative and community-driven process. JMIR Public Health Surveill 2015 Nov 16;1(2):e18 [FREE Full text] [doi: 10.2196/publichealth.4449] [Medline: 27227136]
28. Kenny R, Dooley B, Fitzgerald A. Developing mental health mobile apps: exploring adolescents' perspectives. Health Informatics J 2016 Jun;22(2):265-275 [FREE Full text] [doi: 10.1177/1460458214555041] [Medline: 25385165]
29. Proudfoot J, Parker G, Hadzi Pavlovic D, Manicavasagar V, Adler E, Whitton A. Community attitudes to the appropriation of mobile phones for monitoring and managing depression, anxiety, and stress. J Med Internet Res 2010 Dec 19;12(5):e64 [FREE Full text] [doi: 10.2196/jmir.1475] [Medline: 21169174]
30. Seh AH, Zarour M, Alenezi M, Sarkar AK, Agrawal A, Kumar R, et al. Healthcare data breaches: insights and implications. Healthcare (Basel) 2020 May 13;8(2):133 [FREE Full text] [doi: 10.3390/healthcare8020133] [Medline: 32414183]
31. Bondaronek P, Alkhaldi G, Slee A, Hamilton FL, Murray E. Quality of publicly available physical activity apps: review and content analysis. JMIR Mhealth Uhealth 2018 Mar 21;6(3):e53 [FREE Full text] [doi: 10.2196/mhealth.9069] [Medline: 29563080]
32. Müthing J, Jäschke T, Friedrich CM. Client-focused security assessment of mHealth apps and recommended practices to prevent or mitigate transport security issues. JMIR Mhealth Uhealth 2017 Oct 18;5(10):e147 [FREE Full text] [doi: 10.2196/mhealth.7791] [Medline: 29046271]
33. Papageorgiou A, Strigkos M, Politou E, Alepis E, Solanas A, Patsakis C. Security and privacy analysis of mobile health applications: the alarming state of practice. IEEE Access 2018 Jan 29;6:9390-9403. [doi: 10.1109/access.2018.2799522]
34. Klopfer PH, Rubenstein DI. The concept privacy and its biological basis. J Soc Issues 1977;33(3):52-65. [doi: 10.1111/j.1540-4560.1977.tb01882.x]
35. Posner RA. The economics of privacy. Am Econ Rev 1981 May;71(2):405-409.
36. Norberg PA, Horne DR, Horne DA. The privacy paradox: personal information disclosure intentions versus behaviors. J Consum Aff 2007 Mar 6;41(1):100-126. [doi: 10.1111/j.1745-6606.2006.00070.x]
37. Acquisti A. Privacy in electronic commerce and the economics of immediate gratification. In: Proceedings of the 5th ACM Conference on Electronic commerce. 2004 Presented at: EC '04; May 17-20, 2004; New York, NY, USA p. 21-29. [doi: 10.1145/988772.988777]
38. Hallam C, Zanella G. Online self-disclosure: the privacy paradox explained as a temporally discounted balance between concerns and rewards. Comput Human Behav 2017 Mar;68:217-227. [doi: 10.1016/j.chb.2016.11.033]
39. Hoffman LJ. Computers and privacy: a survey. ACM Comput Surv 1969 Jun;1(2):85-103. [doi: 10.1145/356546.356548]
40. Barth S, de Jong MD. The privacy paradox – investigating discrepancies between expressed privacy concerns and actual online behavior – a systematic literature review. Telemat Inform 2017 Nov;34(7):1038-1058. [doi: 10.1016/j.tele.2017.04.013]
41. Pavlou PA. State of the information privacy literature: where are we now and where should we go? MIS Q 2011 Dec;35(4):977-988. [doi: 10.2307/41409969]
42. Aïmeur E. Personalisation and privacy issues in the age of exposure. In: Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization. 2018 Presented at: UMAP '18; July 8-11, 2018; Singapore, Singapore p. 373-376. [doi: 10.1145/3209219.3209271]
43. Varshney U. Mobile health: four emerging themes of research. Decis Support Syst 2014 Oct;66:20-35. [doi: 10.1016/j.dss.2014.06.001]
44. Xu H, Zhang C, Shi P, Song P. Exploring the role of overt vs. covert personalization strategy in privacy calculus. Acad Manag Proc 2009 Aug 15;2009(1):1-6. [doi: 10.5465/ambpp.2009.44249857]
45. Sutanto J, Palme E, Tan CH, Phang CW. Addressing the personalization-privacy paradox: an empirical assessment from a field experiment on smartphone users. MIS Q 2013 Dec;37(4):1141-1164. [doi: 10.25300/misq/2013/37.4.07]
46. Awad NF, Krishnan MS. The personalization privacy paradox: an empirical evaluation of information transparency and the willingness to be profiled online for personalization. MIS Q 2006 Mar;30(1):13-28. [doi: 10.2307/25148715]
47. Gefen D, Straub DW. Gender differences in the perception and use of e-mail: an extension to the technology acceptance model. MIS Q 1997 Dec;21(4):389-400. [doi: 10.2307/249720]
48. MORRIS MG, VENKATESH V. Age differences in technology adoption decisions: implications for a changing work force. Pers Psychol 2000 Jun;53(2):375-403. [doi: 10.1111/j.1744-6570.2000.tb00206.x]
49. Morris MG, Venkatesh V, Ackerman PL. Gender and age differences in employee decisions about new technology: an extension to the theory of planned behavior. IEEE Trans Eng Manage 2005 Feb;52(1):69-84. [doi: 10.1109/tem.2004.839967]
50. Wang ES. Perceived control and gender difference on the relationship between trialability and intent to play new online games. Comput Human Behav 2014 Jan;30:315-320. [doi: 10.1016/j.chb.2013.09.016]

51. Venkatesh V, Thong JY, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q 2012 Mar;36(1):157-178. [doi: 10.2307/41410412]
52. Featherman MS, Pavlou PA. Predicting e-services adoption: a perceived risk facets perspective. Int J Hum Comput Stud 2003 Oct;59(4):451-474. [doi: 10.1016/s1071-5819(03)00111-3]
53. Reuter T, Ziegelmann JP, Wiedemann AU, Lippke S, Schuz B, Aiken LS. Planning bridges the intention-behaviour gap: age makes a difference and strategy use explains why. Psychol Health 2010 Sep;25(7):873-887. [doi: 10.1080/08870440902939857] [Medline: 20204963]
54. Ziefle M, Röcker C. Acceptance of pervasive healthcare systems: a comparison of different implementation concepts. In: Proceedings of the 4th International Conference on Pervasive Computing Technologies for Healthcare. 2010 Presented at: PERVASIVEHEALTH '10; March 22-25, 2010; Munich, Germany p. 1-6. [doi: 10.4108/ICST.PERVASIVEHEALTH2010.8915]
55. Sintonen S, Immonen M. Telecare services for aging people: assessment of critical factors influencing the adoption intention. Comput Human Behav 2013 Jul;29(4):1307-1317. [doi: 10.1016/j.chb.2013.01.037]
56. Guo X, Sun Y, Wang N, Peng Z, Yan Z. The dark side of elderly acceptance of preventive mobile health services in China. Electron Mark 2012 Dec 11;23(1):49-61. [doi: 10.1007/s12525-012-0112-4]
57. Miles MB, Huberman M, Saldana J. Qualitative Data Analysis: A Methods Sourcebook. 3rd edition. Thousand Oaks, CA, USA: Sage Publications; 2014.
58. Flick U. An Introduction to Qualitative Research. 6th edition. Thousand Oaks, CA, USA: Sage Publications; Jan 2019.
59. Thomas DR. A general inductive approach for analyzing qualitative evaluation data. Am J Eval 2006 Jun 1;27(2):237-246. [doi: 10.1177/1098214005283748]
60. Swoboda W, Schmieder M, Bulitta C, Gaisser S, Hofmann GR, Kremer-Rücker P, et al. Die gemeinsame Ethik-Kommission der Hochschulen Bayerns – GEHBa. mdi - Forum der Medizin-Dokumentation und Medizin-Informatik 2021;3:80-83 [FREE Full text]
61. Xu H, Dinev T, Smith HF, Hart P. Examining the formation of individual's privacy concerns: toward an integrative view. In: Proceedings of the 2008 International Conference on Information Systems. 2008 Presented at: ICIS '08; December 14-17, 2008; Paris, France p. 6.
62. Baruh L, Secinti E, Cemalcilar Z. Online privacy concerns and privacy management: a meta-analytical review. J Commun 2017 Jan 17;67(1):26-53. [doi: 10.1111/jcom.12276]
63. Fruchter N, Liccardi I. Consumer attitudes towards privacy and security in home assistants. In: Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI EA '18; April 21-26, 2018; Montreal, Canada p. 1-6. [doi: 10.1145/3170427.3188448]
64. Kennedy S, Li H, Wang C, Liu H, Wang B, Sun W. I can hear your Alexa: voice command fingerprinting on smart home speakers. In: Proceedings of the 2019 IEEE Conference on Communications and Network Security. 2019 Presented at: CNS '19; June 10-12, 2019; Washington, DC, USA p. 232-240. [doi: 10.1109/cns.2019.8802686]
65. Gross R, Acquisti A. Information revelation and privacy in online social networks. In: Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society. 2005 Presented at: WPES '05; November 7, 2005; Alexandria, VA, USA p. 71-80. [doi: 10.1145/1102199.1102214]
66. Kokolakis S. Privacy attitudes and privacy behaviour: a review of current research on the privacy paradox phenomenon. Comput Secur 2017 Jan;64:122-134. [doi: 10.1016/j.cose.2015.07.002]
67. Smith HJ, Dinev T, Xu H. Information privacy research: an interdisciplinary review. MIS Q 2011 Dec;35(4):989-1015. [doi: 10.2307/41409970]
68. Bansal G, Zahedi FM, Gefen D. The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online. Decis Support Syst 2010 May;49(2):138-150. [doi: 10.1016/j.dss.2010.01.010]
69. Bhattacherjee A, Premkumar G. Understanding changes in belief and attitude toward information technology usage: a theoretical model and longitudinal test. MIS Q 2004 Jun;28(2):229-254. [doi: 10.2307/25148634]
70. Sun H. A longitudinal study of herd behavior in the adoption and continued use of technology. MIS Q 2013 Dec;37(4):1013-1041. [doi: 10.25300/misq/2013/37.4.02]
71. Jiang R. Introducing new Alexa healthcare skills. Amazon Alexa. 2019 Apr 4. URL: https://developer.amazon.com/de/blogs/alexa/post/ff33dbc7-6cf5-4db8-b203-99144a251a21/introducing-new-alexa-healthcare-skills [accessed 2021-06-25]
72. Koppel R, Kuziemsky C. Healthcare data are remarkably vulnerable to hacking: connected healthcare delivery increases the risks. Stud Health Technol Inform 2019;257:218-222. [Medline: 30741199]
73. Lee DJ, Ahn JH, Bang Y. Managing consumer privacy concerns in personalization: a strategic analysis of privacy protection. MIS Q 2011 Jun;35(2):423-444. [doi: 10.2307/23044050]

Abbreviations

mHealth: mobile health
RQ: research question

Edited by G Eysenbach; submitted 17.02.21; peer-reviewed by E Baker, D Kotz, I Schiering; comments to author 24.03.21; revised version received 30.06.21; accepted 16.04.22; published 13.06.22.

Please cite as:
Schroeder T, Haug M, Gewald H
Data Privacy Concerns Using mHealth Apps and Smart Speakers: Comparative Interview Study Among Mature Adults
JMIR Form Res 2022;6(6):e28025
URL: https://formative.jmir.org/2022/6/e28025
doi:10.2196/28025
PMID:35699993

Original Paper

Applying the Health Belief Model to Characterize Racial/Ethnic Differences in Digital Conversations Related to Depression Pre- and Mid-COVID-19: Descriptive Analysis

Ruby Castilla-Puentes[1,2,3], MBA, MD, DrPH; Jacqueline Pesa[4], MSED, MPH, PhD; Caroline Brethenoux[5], MS; Patrick Furey[5], MS; Liliana Gil Valletta[5], MBA; Tatiana Falcone[6], MPH, MD

[1]Janssen Research & Development, LLC, Titusville, NJ, United States

[2]Center for Public Health Practice, Drexel University, Philadelphia, PA, United States

[3]Hispanic Organization for Leadership and Advancement, Johnson & Johnson, Employee Resource Group, New Brunswick, NJ, United States

[4]Janssen Scientific Affairs, LLC, Titusville, NJ, United States

[5]CulturIntel, Inc, New York, NY, United States

[6]Department of Psychiatry and Psychology, Cleveland Clinic Lerner College of Medicine, Cleveland, OH, United States

Corresponding Author:
Ruby Castilla-Puentes, MBA, MD, DrPH
Center for Public Health Practice
Drexel University
530 S 2nd st Suite 743
Philadelphia, PA, 19147
United States
Phone: 1 6108642528
Email: rcastil4@its.jnj.com

Abstract

Background: The prevalence of depression in the United States is >3 times higher mid-COVID-19 versus prepandemic. Racial/ethnic differences in mindsets around depression and the potential impact of the COVID-19 pandemic are not well characterized.

Objective: This study aims to describe attitudes, mindsets, key drivers, and barriers related to depression pre- and mid-COVID-19 by race/ethnicity using digital conversations about depression mapped to health belief model (HBM) concepts.

Methods: Advanced search, data extraction, and artificial intelligence–powered tools were used to harvest, mine, and structure open-source digital conversations of US adults who engaged in conversations about depression pre- (February 1, 2019-February 29, 2020) and mid-COVID-19 pandemic (March 1, 2020-November 1, 2020) across the internet. Natural language processing, text analytics, and social data mining were used to categorize conversations that included a self-identifier into racial/ethnic groups. Conversations were mapped to HBM concepts (ie, perceived susceptibility, perceived severity, perceived benefits, perceived barriers, cues to action, and self-efficacy). Results are descriptive in nature.

Results: Of 2.9 and 1.3 million relevant digital conversations pre- and mid-COVID-19, race/ethnicity was determined among 1.8 million (62.2%) and 979,000 (75.3%) conversations, respectively. Pre-COVID-19, 1.3 million (72.1%) conversations about depression were analyzed among non-Hispanic Whites (NHW), 227,200 (12.6%) among Black Americans (BA), 189,200 (10.5%) among Hispanics, and 86,800 (4.8%) among Asian Americans (AS). Mid-COVID-19, a total of 736,100 (75.2%) conversations about depression were analyzed among NHW, 131,800 (13.5%) among BA, 78,300 (8.0%) among Hispanics, and 32,800 (3.3%) among AS. Conversations among all racial/ethnic groups had a negative tone, which increased pre- to mid-COVID-19; finding support from others was seen as a benefit among most groups. Hispanics had the highest rate of any racial/ethnic group of conversations showing an avoiding mindset toward their depression. Conversations related to external barriers to seeking treatment (eg, stigma, lack of support, and lack of resources) were generally more prevalent among Hispanics, BA, and AS than among NHW. Being able to benefit others and building a support system were key drivers to seeking help or treatment for all racial/ethnic groups.

Conclusions: There were considerable racial/ethnic differences in drivers and barriers to seeking help and treatment for depression pre- and mid-COVID-19. As expected, COVID-19 has made conversations about depression more negative and with frequent

discussions of barriers to seeking care. Applying concepts of the HBM to data on digital conversation about depression allowed organization of the most frequent themes by race/ethnicity. Individuals of all groups came online to discuss their depression. These data highlight opportunities for culturally competent and targeted approaches to addressing areas amenable to change that might impact the ability of people to ask for or receive mental health help, such as the constructs that comprise the HBM.

(JMIR Form Res 2022;6(6):e33637) doi:10.2196/33637

KEYWORDS

depression; COVID-19; treatment; race/ethnicity; digital conversations; health belief model; artificial intelligence; natural language processing

Introduction

Worldwide, depression is the leading cause of years lost due to disability and is associated with excess mortality [1]. Data from the 2019 National Health Interview Survey showed that 18.5% of US adults had symptoms of depression, which can include depressed mood, lack of interest or pleasure in daily activities, weight loss, sleep disturbances, psychomotor issues, lack of energy, feelings of worthlessness or guilt, difficulty concentrating, and suicidal thoughts or actions, in the preceding 2 weeks [2,3].

The same survey data indicated that adult non-Hispanic Whites (NHW) and non-Hispanic Blacks were the most likely to have experienced depression symptoms in the preceding 2 weeks (19.3% each), followed by Hispanic adults (16.9%) and Asian adults (10.2%) [2]. Furthermore, data suggest that the prevalence of depression symptoms in the United States was more than threefold higher during the COVID-19 pandemic than before the pandemic [4]. Despite the significant prevalence of depression symptoms across racial/ethnic groups, the likelihood of receiving treatment for depression is significantly lower for Black Americans (BA), Hispanics, and Asian Americans (AS) compared with NHW [5,6]. Prior to the COVID-19 pandemic, barriers existed that prevented these ethnic groups from seeking mental health care, which contributed to the lower likelihood of receiving treatment. Such barriers include an increased stigma with regard to experiencing depressive symptoms, a lack of education/health literacy, cultural and language barriers, not being able to access a health care professional (HCP) because of the inability to leave or miss work, or prior experience with mistreatment/misdiagnosis by an HCP [5,7-17]. Additionally, individuals in these groups may be fearful of seeking treatment for their depression if they believe they will be met with racism or threats of deportation [12,13,18]. As the pandemic continues, the specific effects of COVID-19 on drivers and previously existing barriers to seeking help or treatment for depression are not well understood.

The health belief model (HBM) is a theoretical framework that can be used to better understand help- and treatment-seeking behaviors and has previously been used among those with depression [19,20]. The HBM considers the value one assigns to maintaining wellness or seeking treatment in the face of an illness and one's beliefs about the effect of taking action [21]. It posits that behavior can be understood when the value an individual places on a particular outcome is known, as well as the likelihood (ie, expectation) that the action would result in the desired outcome [21].

Despite recent research efforts, there is a lack of information on how people of different racial/ethnic groups perceive and seek help or treatment for depression and how this has been impacted by the COVID-19 pandemic. Information gathered from conversations that take place online can provide new insights into how people are impacted by their disease, how to target interventions to decrease barriers to care, and how to help patients and families be more comfortable seeking and receiving mental health care.

This study applied big data and artificial intelligence (AI) techniques to analyze open-source digital conversations pre- and mid-COVID-19 by race/ethnicity to identify differences in attitudes, mindsets, key drivers, and barriers to depression care. Efforts were made to map the resulting conversations to HBM concepts in order to guide targeted outreach to specific racial/ethnic groups and inform culturally competent interventions to address depression in these communities. Such efforts have the potential to narrow gaps in known disparities with respect to mental health care.

Methods

Data Source

CulturIntel harvested, mined, and structured open-source digital conversations in both English and Spanish from adults who engaged in conversation about depression pre- (February 1, 2019-February 29, 2020) and mid-COVID-19 (March 1, 2020-November 1, 2020).

Only digital conversations originating from US internet protocol addresses were analyzed. Sources of conversations included message boards (ie, any online discussion site where people can hold conversations in the form of posted messages), topical sites (ie, any site that relates to a specific topic, in this case mental health and depression), social networks (eg, Facebook, Twitter, and Instagram), content-sharing sites (eg, YouTube), blogs (ie, any regularly updated website or web page, typically 1 run by an individual or a small group), and comments (ie, any mention related to mental health or depression posted publicly in an open comment box).

Rather than using keywords, conversations were mined by the topic of depression. Discussions were identified and included if they were related to depression in general (defined by the use of the term "depression" and its adjacencies, such as "feeling depressed"), seeking help for depression (defined by the use of terms such as "help," "looking for," "support," and "assistance"), and depression and COVID-19 (defined by including "depression" and "COVID-19" terms). Each unique

comment across discussions/posts/sites was counted for this analysis. A single comment, if appearing repeatedly through sharing or linking, was counted and analyzed once.

Conversant Demographics and Characteristics

After the completion of comprehensive data collection, topical and tags data were extracted using the CulturIntel methodology, with the origin and user criteria based on self-identification, and a large, unstructured "big" data set was created. Demographic characteristics were determined by scanning user profiles for self-entered data and by scanning within the text for self-identification. Demographic characteristics and comorbidities related to conversations were captured in order to better profile the population behind the conversations. To categorize conversations by racial/ethnic group, the conversations had to include a self-identifier: NHW, BA, Hispanic, or AS. The clustering and tagging of the conversations as pertaining to a specific group either defined by demographic characteristics or comorbidities was based on self-identification (ie, how people self-identify in the conversation itself or on their public profile).

Most national government agencies, such as the US Census Bureau, and many research organizations use the descriptors "Hispanics," "Latinos," and "Latinx" interchangeably as umbrella terms for Hispanic Americans. It is possible that members of this population may elect to self-identify as 1 or many of these identities concurrently. In this study, the descriptor "Hispanic(s)" was used to describe the ethnicity of individuals who identify as any of the above-mentioned racial/ethnic groups. Similarly, the descriptor "BA" was used to describe the ethnicity of individuals who identify as "African" or "African American," "NHW" for those who identify as "White" without any mention of other race or ethnicity, "Caucasian," or "of European descent," and "AS" to describe ethnicities of Asian Pacific descent.

Where possible, conversations among those with COVID-19 or those with a loved one with COVID-19 were identified using the same self-identification methodology.

Thematic Analysis and HBM Mapping

Natural language processing (NLP) is a subfield of AI that helps computers to process and analyze large amounts of natural human language data and is broadly considered to be the study and development of computer systems that can interpret speech and text as humans naturally speak and type it. Text analytics refers to the computer-based processes used for deriving high-quality information from text. Machine learning for NLP and text analytics involves a set of statistical algorithms and rules for identifying and recognizing parts of speech, named entities, sentiment, themes, and other aspects of text. Low-level text functions are the first processes through which any text is initially analyzed, and can include tokenization (breaking text into data, such as words or groups of words), part-of-speech tagging (identifying nouns, adverbs, adjectives of each token), named entity recognition (pretagged entities), sentiment analysis (whether data is positive, negative, or neutral and devising weighted sentiment scores), and sentence boundaries and syntax analysis [22-24]. Midlevel text functions involve extracting the real content of text, including entities, themes, topics, summaries, and intentions. Finally, a high-level text function is the application of sentiment to the text. This study used a linear support vector machines model with a classification accuracy of 90.6% [25]. Using machine learning and predefined rules, text data were tagged or annotated with examples of what the machine should look for and how it should interpret that aspect. These data were used as a training set for a statistical model, which was then given untagged text to analyze. In this study, these analyses were human assisted and included repeated training, testing, and reviewing of the program output by the CulturIntel team and the authors.

These methods were used in conjunction with social media data mining to examine patterns in data and perform a thematic analysis [23,24,26]. For the thematic analysis, CulturIntel tagged and sorted data; determined key sentiments toward depression, drivers of those sentiments, motivations and barriers to seeking help, and overarching mindsets; and assigned underlying drivers and barriers, when possible, throughout decision journey stages. Conversations in English and Spanish were analyzed together.

Components of conversations were mapped to HBM concepts to explore and understand attitudes and mindsets toward depression and key drivers and barriers to seeking help and treatment among different racial/ethnic groups (Table 1). The HBM is composed of 6 constructs that predict an individual's readiness to enact change [21]. Perceived susceptibility describes an individual's belief that they are at risk for the health problem or associated negative outcome. Perceived severity relates to an individual's feelings on the seriousness of contracting an illness or disease (or leaving the illness or disease untreated). Perceived benefits refer to an individual's perception of the effectiveness of various actions available to reduce the threat of illness or disease (or to cure illness or disease). Perceived barriers are an individual's feelings on the obstacles to performing a recommended health action; the person weighs the effectiveness of the actions against the perceptions that it may be expensive, dangerous (eg, side effects), unpleasant (eg, painful), time-consuming, or inconvenient. Cues to action refer to the stimuli (internal or external) necessary to trigger the decision to engage in a behavior. Finally, self-efficacy is the level of an individual's confidence in their ability to successfully perform a behavior.

Through NLP, the tone of the sentiment (ie, negative, neutral, or positive) of the conversations about depression was categorized and used as a proxy for individuals' attitudes toward seeking help or treatment. A negative sentiment toward depression was mapped to the HBM construct of perceived susceptibility. Perceived severity was mapped to drivers of negative sentiments toward the self, the future, and the world (ie, losing the quality of life, being unable to function, feeling like a burden, feeling hopeless, dealing with uncertainty, feeling stigmatized, or feeling a lack of support). Drivers of a negative sentiment toward seeking help were mapped to the HBM construct of perceived barriers; these related to the future and the world (ie, feeling helpless, feeling stigmatized, lacking resources, facing barriers, lacking support, or dealing with misinformation). Drivers of a positive sentiment toward depression related to the self, the future, and the world (ie,

improving the quality of life, making progress, having a sense of agency, finding support, or getting access to treatment) were mapped to the HBM construct of perceived benefit. A neutral sentiment toward depression (ie, seeking support and treatment, understanding their situation, or looking for information and resources) was mapped to the HBM construct of cues to action; positive drivers for seeking help related to the future and the world (ie, benefiting others, building a strong support system, being encouraged by an HCP, having access to effective help, or getting helpful knowledge) were also mapped to the cues to action construct.

A mindset is a set of beliefs that orients the way a person handles their depression. Through NLP, conversations about depression reflecting individuals' mindsets toward treatment were classified as 1 of 4 possible mindsets. The denial and troubled mindsets make up the avoiding approach, and the overcome and empowered mindsets make up the confrontational approach. The denial mindset involves ignoring the condition and declining to actively seek treatment in order to tolerate it (eg, struggling with depression but putting it aside in order to continue to care for the family). The troubled mindset involves acknowledgement of the condition but uncertainty with regard to the next steps to take or a feeling of hopelessness with regard to improvement (eg, feeling depressed, and doubtful it will get better). The overcome mindset involves reinforcing a positive outlook (eg, after being diagnosed, realizing I will get better). The empowered mindset involves recognizing that depression is an illness and taking action to feel better (eg, although depression is challenging, feeling it is within my power to cope with my symptoms and improve my quality of life). The denial and troubled mindsets were mapped to the HBM concept of perceived barriers, and the empowered mindset was mapped to the HBM concept of self-efficacy.

Table 1. Mapping conversations to HBM[a] concepts.

HBM concept	Sentiment analysis
Perceived susceptibility	• Negative sentiment levels toward depression
Perceived severity	• Drivers of a negative sentiment toward depression
Perceived benefits	• Drivers of a positive sentiment toward depression
Perceived barriers	• Denial and troubled mindsets along the path to treatment • Drivers of a negative sentiment toward seeking help for depression
Cues to action	• Neutral sentiment level toward depression • Drivers of seeking help for depression
Self-efficacy	• Empowered mindset along the path to treatment

[a]HBM: health belief model.

Numerical Analyses

Characteristics of the overall population and key segments were described by the number of conversations among subgroups of interest. Results are descriptive in nature, and no formal statistical analyses were performed.

Ethics Consideration

All the information gathered from the different online, open sources (topical sites [eg, Depression and Bipolar Support Alliance], message boards [eg, Beyond Blue], social networks [eg, Facebook], and blogs) is in the public domain and is deidentified. The study was exempt from Institutional Review Board approval as it used publicly available, deidentified information.

Results

Digital Conversations About Depression

A total of 4.2 million relevant digital conversations about depression occurring between February 2019 and November 2020 were analyzed: 2.9 million (69.1%) in the 12-month pre-COVID-19 period (February 1, 2019-February 29, 2020) and 1.3 million (30.9%) in the mid-COVID-19 period (March 1, 2020-November 1, 2020). The majority of conversations both pre- and mid-COVID-19 took place on topical sites (n=1,102,000 conversations [38%] and n=507,000 conversations [39%], respectively) or message boards (n=754,000 conversations [26%] and n=377,000 conversations [29%], respectively; Figure 1).

Of the 2.9 million pre-COVID-19 conversations about depression, race/ethnicity was determined in 1.8 million (62.2%) conversations; 1.3 million (72.1%) conversations occurred among NHW, 227,200 (12.6%) among BA, 189,200 (10.5%) among Hispanics, and 86,800 (4.8%) among AS. Of the 1.3 million conversations mid-COVID-19, race/ethnicity was determined in 979,000 (75.3%) conversations; 736,100 (75.2%) conversations about depression occurred among NHW, 131,800 (13.5%) among BA, 78,300 (8.0%) among Hispanics, and 32,800 (3.3%) among AS (Figure 2). Mid-COVID-19, a greater proportion of conversations about depression occurred among those who mentioned having COVID-19 or having a loved one with COVID-19 among Hispanics (n=26,100 conversations, 33.3%) and BA (n=43,800 conversations, 33.2%) than AS (n=7900 conversations, 24.1%) and NHW (n=157,800 conversations, 21.4%).

Figure 1. Sources of relevant digital conversations about depression pre- and mid-COVID-19.

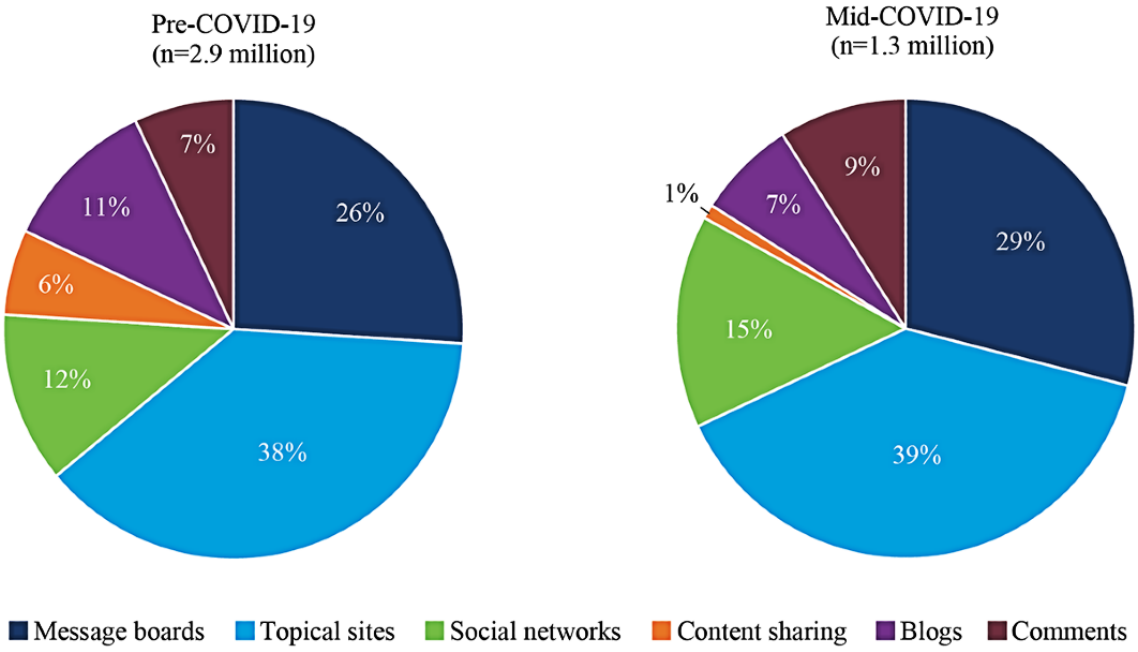
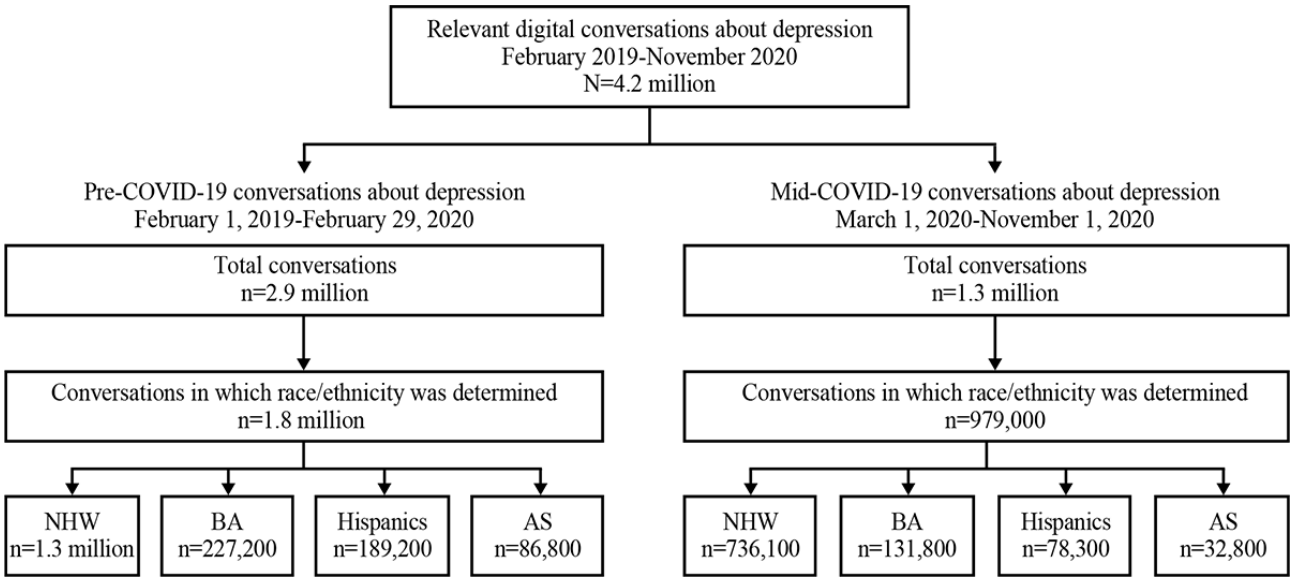


Figure 2. Summary of relevant digital conversations about depression pre- and mid-COVID-19. AS: Asian Americans; BA: Black Americans; NHW: non-Hispanic Whites.



Perceived Susceptibility and Perceived Severity

In the HBM, perceived susceptibility is driven by an individual's belief that they are at risk for negative outcomes and perceived severity, which relates to beliefs about whether the problem is serious enough to warrant treatment. In this study, both perceived susceptibility and perceived severity were characterized by negative sentiment and the key themes that can be considered drivers. Across racial/ethnic groups, conversations about depression were predominantly negative and Hispanics had the highest proportion of conversations with a negative sentiment (Figure 3). The proportion of conversations that expressed a negative sentiment toward depression increased from pre- to mid-COVID-19 across most racial/ethnic groups, with the greatest increase observed among AS (52,080/86,800 [60%] to 24,928/32,800 [76%]). Mid-COVID-19, conversations among BA, Hispanics, and AS all had greater proportions of negative sentiment compared with NHW.

For all groups except Hispanics, losing the quality of life was the dominant topic of conversations, with substantial changes pre- to mid-COVID-19. An inability to function also emerged as a concern for all groups (Figure 4A). Feeling like a burden to others was expressed more frequently for Hispanics at both time points than for other groups. More conversations among Hispanics and BA involved stigma around depression compared with NHW and AS both pre- and mid-COVID-19. Furthermore, a higher proportion of conversations mid-COVID-19 among Hispanics, BA, and AS were related to feeling a lack of support from others (16,327/56,300 [29%], 19,779/104,100 [19%], and 3500/25,000 [14%], respectively) compared with NHW (34,510/493,000 [7%]).

Figure 3. Sentiment by racial/ethnic group pre- and mid-COVID-19. AS: Asian Americans; BA: Black Americans; NHW: non-Hispanic Whites.

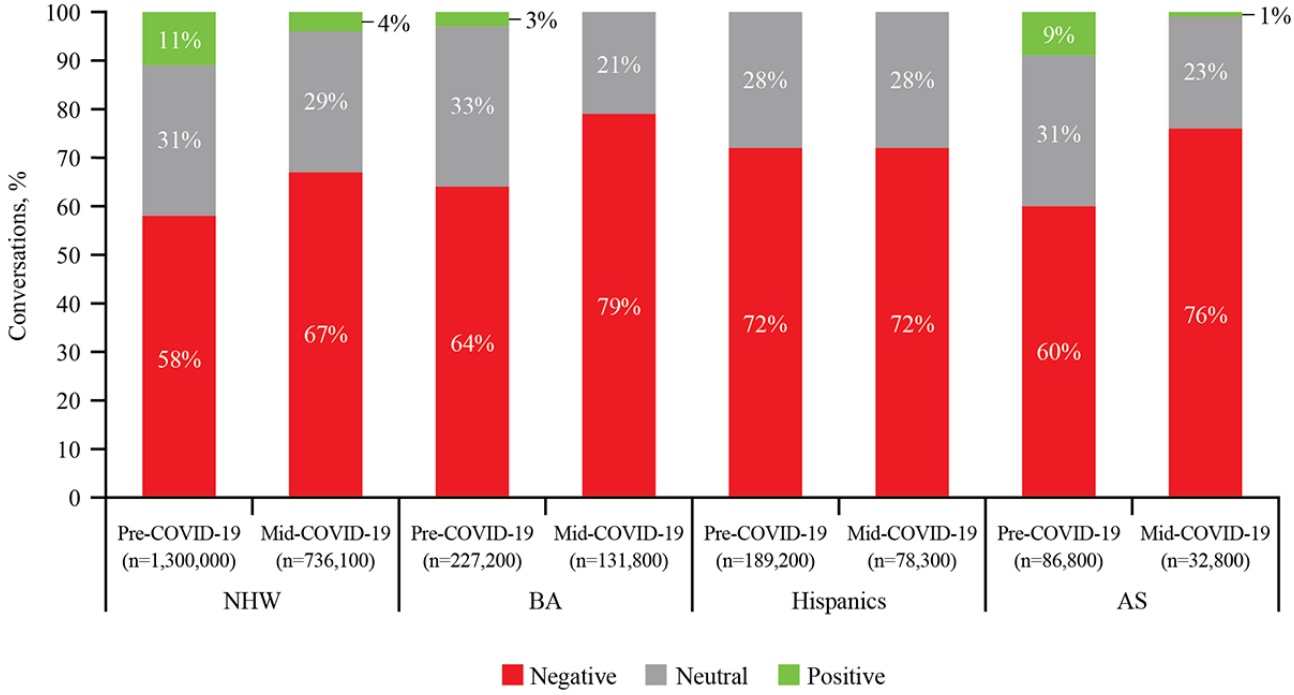
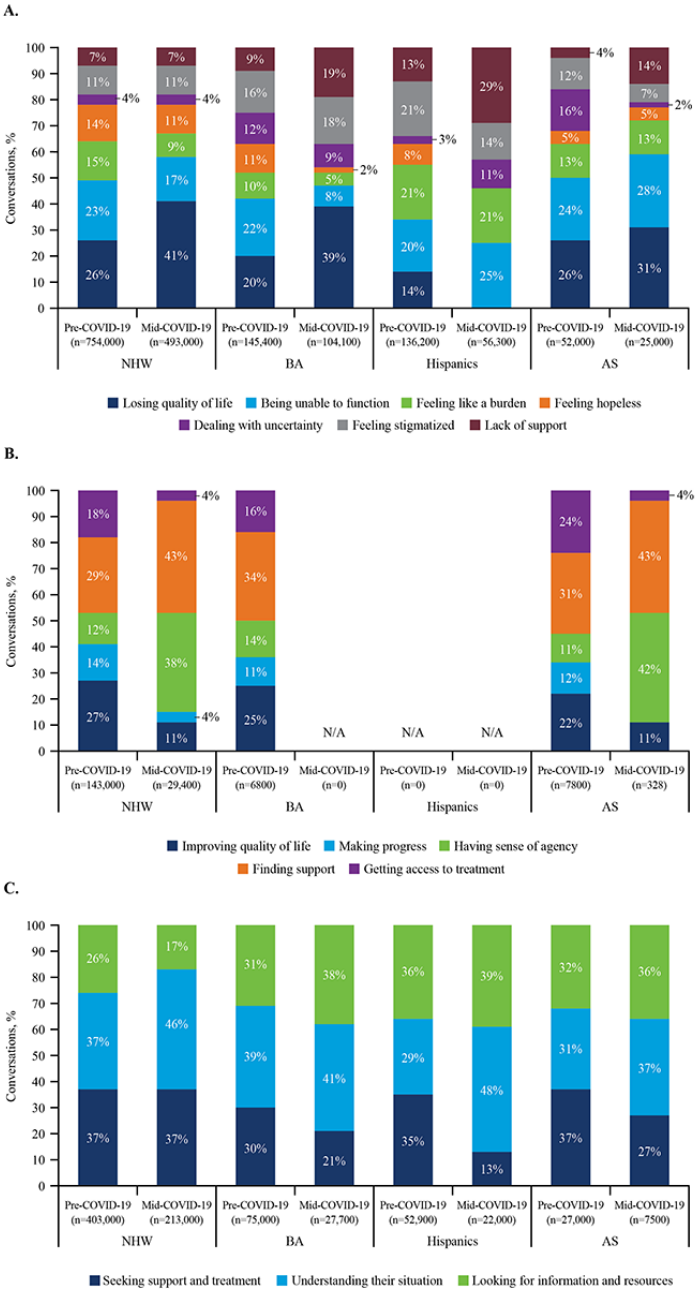


Figure 4. Drivers of (A) negative, (B) positive, and (C) neutral sentiment by racial/ethnic group among conversations about depression pre- and mid-COVID-19. AS: Asian Americans; BA: Black Americans; NHW: non-Hispanic Whites.



Perceived Benefits

Based on the HBM, perceived benefits relate to beliefs about the benefits of treatment; perceived benefits were mapped to drivers of a positive sentiment toward depression. Conversations revealed little positive sentiment toward depression across racial/ethnic groups. Further, a positive sentiment among NHW, BA, and AS decreased from pre- to mid-COVID-19; Hispanics displayed no positive sentiment toward depression both pre- and mid-COVID-19.

Finding support was important across the NHW, BA, and AS groups, as evidenced by the highest frequency of conversations (Figure 4B). In the NHW group pre-COVID-19, a positive sentiment toward depression was most often expressed by finding support (41,470/143,000 [29%]) and improving the quality of life (38,610/143,00 [27%]). Mid-COVID-19, finding support was discussed even more often (12,642/29,400 [43%]), and having a sense of agency seemed to become more relevant than it had been pre-COVID-19. In conversations among BA pre-COVID-19, the small amount of positive sentiment was driven by finding support (2312/6800 [34%]); a positive sentiment toward depression was absent in conversations among BA captured mid-COVID-19. Conversations in the AS subpopulation pre-COVID-19 were primarily dominated by finding support (2418/7800 [31%]) and getting access to treatment (1872/7800 [24%]). Mid-COVID-19, a positive sentiment among AS significantly decreased, but the remaining positive sentiment was tied to finding support (141/328 [43%]) and having a sense of agency (138/328 [42%]).

Perceived Barriers

The HBM concept of perceived barriers, which is related to beliefs about internal/external barriers to treatment, was mapped to the mindset of avoiding accessing treatment for their depression and drivers of a negative sentiment toward seeking help or treatment for depression. The way a person handles their depression is related to their mindset, which can range from avoiding to confronting. Avoiding mindsets include denial (ie, ignoring the condition to tolerate it) and troubled (ie, carrying the burden without acting), while confronting mindsets include overcome (ie, reinforcing a positive mindset) and empowered (ie, taking action to feel better).

Pre-COVID-19 conversations indicate that, of all groups, Hispanics had the most avoiding mindsets (112,000/130,233, 86% of conversations), followed by BA (156,000/232,825, 67% of conversations), NHW (396,000/707,143, 53% of conversations), and AS (37,000/77,083, 48% of conversations); these mindsets were not substantially different mid-COVID-19.

All racial/ethnic groups faced external barriers to getting help for depression. Stigma was a significant driver of negative sentiment in all groups seeking help for depression both pre- and mid-COVID-19. Pre- and mid-COVID-19, BA, Hispanics, and AS all faced more external practical and logistical barriers to getting help (eg, lack of transportation/long commute to the point of care, inability to secure adequate time off or child/elderly care to engage with care) than NHW; among those seeking help for depression mid-COVID-19, 36% (5000/13,878), 31% (4000/12,903), and 29% (1000/3447) of conversations among BA, Hispanics, and AS, respectively, expressed concerns about external barriers compared with 17% (17,000/99,989) of conversations among NHW.

Of all groups, Hispanics most frequently expressed that they did not have access to a support system (mid-COVID-19, 16,327/56,300 [29%] of barriers to seeking help; Figure 4A). Hispanics also mentioned lacking resources for getting help more frequently than other racial/ethnic groups (mid-COVID-19, 23% [8000/34,781] of barriers to getting help). BA also often felt they lacked resources for getting help (mid-COVID-19, 18% [17000/94,442] of barriers to getting help).

Cues to Action

Based on the HBM, cues to action are factors that increase treatment readiness, and these were mapped to neutral sentiments toward depression and drivers for seeking help for depression. The proportion of conversations expressing neutral sentiments toward depression pre- and mid-COVID-19 were relatively consistent among NHW and Hispanics, while conversations with a neutral sentiment decreased among BA and AS, consistent with increasingly negative conversations about depression. Conversations about understanding their situation among Hispanics increased from 15,341 (29%) of 52,900 conversations to 10,560 (48%) of 22,000 conversations pre- to mid-COVID-19 (Figure 4C).

Compared with NHW, Hispanics, BA, and AS had more conversations about looking for information and resources pre-COVID-19 (104,780/403,000 [26%], 19,044/52,900, [36%], 23,250/75,000 [31%], and 8640/27,000 [32%], respectively). Mid-COVID-19, a similar pattern was observed, with a higher proportion of conversations among Hispanics, BA, and AS related to looking for information and resources compared to NHW. There were also fewer conversations among Hispanics, BA, and AS (2860/22,000 [13%], 5817/27,700 [21%], and 2025/7500 [27%], respectively) mid-COVID-19 about seeking support and treatment compared to NHW (78,810/213,000 [37%]).

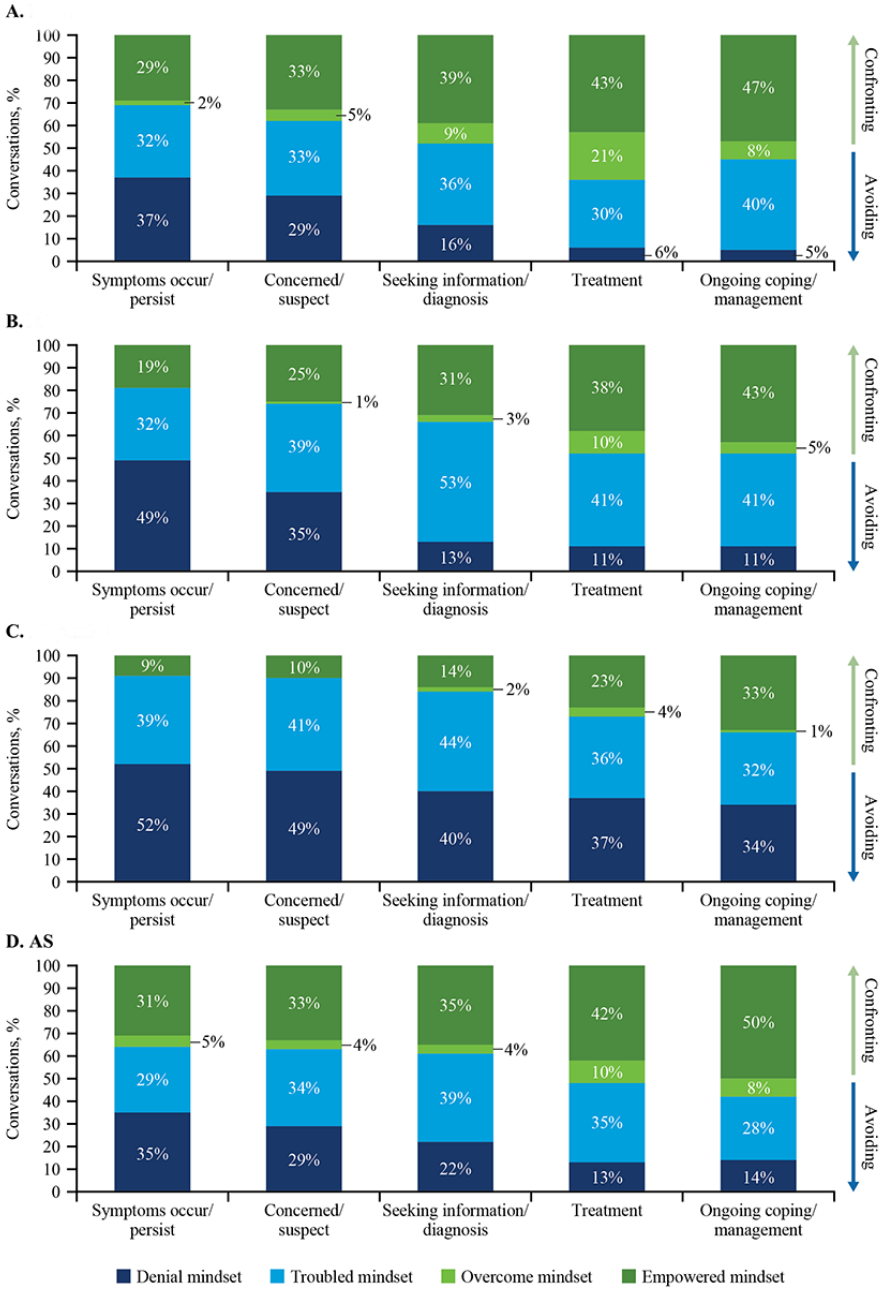
Among those seeking help for depression, being able to benefit others was a driver among all racial/ethnic groups pre-COVID-19 (less than 25% of drivers [121,000/484,000]) but was not a key driver mid-COVID-19 (less than 2% of drivers [7000/350,000]). Among Hispanics, building a strong support system and being encouraged by an HCP were key drivers to seeking help pre-COVID-19 (29% [4000/13,793] and 22% [6000/27,281] of drivers, respectively), but all positive sentiment toward depression had disappeared mid-COVID-19.

Self-efficacy

Based on the HBM, self-efficacy, which relates to how well or poorly a person is able to cope with a given situation based on the skills they have and the circumstances they face, was mapped to having an empowered mindset about depression in this study. Of all groups, conversations among AS were most likely to indicate an empowered mindset about depression (3000/7317, 41% of conversations), followed by NHW (7000/20,588, 34%), BA (4000/13,793, 29%), and Hispanics (2000/14,285, 14%).

Conversations were also examined to determine the mindsets along the path to treatment. The path to treatment starts with the occurrence or persistence of symptoms, then concern or suspicion that the individual may have depression, seeking information or a diagnosis, treatment, and finally ongoing coping and depression management. Along the path to treatment, NHW and AS displayed the highest levels of empowered mindsets (Figure 5). Conversations indicate that few Hispanics started the path to treatment with an empowered mindset (3000/33,327, 9%). Hispanics tended to display the most empowered mindset during the ongoing management stage (4000/12,125, 33%), though it was still lower proportionately than among NHW (11,000/23,404, 47%). Conversations indicate that few in the BA group started the path to treatment with an empowered mindset (4000/21,052, 19%). Across all racial/ethnic groups, the highest level of empowered mindset was observed during the ongoing management phase, suggesting that there is an opportunity to drive adherence to treatment during this phase across racial/ethnic groups.

Figure 5. Mid-COVID-19 conversations about mindsets along the path to treatment from conversations among (A) Non-Hispanic Whites (NHW), (B) Black Americans (BA), (C) Hispanics, and (D) Asian Americans (AS).



Discussion

Principal Findings

This study identified online discussions about depression among individuals of all racial/ethnic groups. There were considerable differences between racial/ethnic groups in drivers and barriers to seeking help and treatment for depression pre- and mid-COVID-19. Generally, conversations about depression were more negative mid-COVID-19, with frequent discussions of barriers to seeking care.

The HBM is a value expectancy theory that has been previously used to better understand help- and treatment-seeking behaviors among those with depression [19,20]. All domains of the HBM have been associated with the likelihood that an individual is receiving treatment for depression [27]. Consistent with previous findings, this study suggests that interventions targeting different HBM concepts may have varying impacts on different ethnic groups [27].

Recently, social media data have also been applied to the COVID-19 pandemic to examine topics ranging from trends in discourse to gender differences in mental health outcomes during the pandemic [28,29], though the effects of the COVID-19 pandemic on depression among different racial/ethnic groups remained unexplored. This study further highlights the utility of online conversations for offering insight into how individuals think about and cope with their disease, including those dealing with mental health issues, such as depression [30-32]. To the best of the authors' knowledge, this study is the first to apply HBM concepts to online conversations to provide new insights into attitudes, mindsets, and help- and treatment-seeking

behaviors related to depression among different racial/ethnic groups pre- and mid-COVID-19.

In this study, the majority of conversations among all racial/ethnic groups discussing their depression online had a negative tone. Pre-COVID-19, the highest proportion of negative sentiment was observed among Hispanics, and the high rate of negative sentiment among Hispanics continued mid-COVID-19. A negative sentiment toward depression increased pre- to mid-COVID-19 in other racial/ethnic groups. Though a relatively small proportion of conversations expressed a positive sentiment toward depression, finding support from others was seen as a benefit for most groups. A previous study conducted among African Americans has shown that partnering with community members is a key factor that can facilitate engagement with health care services [33]. Therefore, efforts by community leaders and HCPs to emphasize the benefits of receiving support from one's own community could help encourage those affected by depression to address their condition.

All racial/ethnic groups who discussed depression online had already seen some negative impacts of depression on their life. For all racial/ethnic groups, perceived barriers to treatment were both internal and external. Internal barriers were mindset driven, and the Hispanic group demonstrated the greatest unwillingness to acknowledge their depressive symptoms; these mindsets did not change with COVID-19. This aligns with previously published studies highlighting the internal stigma with regard to depressive symptoms within the Hispanic community [8,13,16,34]. External barriers to seeking treatment for depression included stigma, lack of support from others, and lack of resources. Stigma and a lack of support from others were particularly relevant for Hispanics and BA, aligning with previous studies that demonstrated how depression symptoms are perceived and responded to among members of these racial/ethnic groups [5,7-9,12,13,16,34]. To effectively help individuals overcome stigma about depression, HCPs should know how to interact with individuals with depression, including the use of person-first behavior; learn about interventions targeting unconscious biases and false beliefs; and understand how they can have an impact on their patients' recovery [35]. Hispanics, BA, and AS were all more likely to face external barriers to treatment compared to NHW. The barriers to health care observed in this study should be considered in the context of racial/ethnic disparities in health care that have existed and continue to exist throughout the United States [5].

Being able to benefit others was a key driver to seeking help pre-COVID-19 for all racial/ethnic groups; thus, leveraging loved ones and positioning treatment as a way to help those suffering from depression offer and receive support from loved ones are key opportunities for HCPs to trigger action. Furthermore, building a strong support system was the most prominent driver to seeking help for many groups. Thus, mid-COVID-19, creating a support system while observing and promoting social distancing is critical.

Of all racial/ethnic groups, the AS group had the most empowered attitude toward depression. However, once individuals reached the ongoing management stage of the path to treatment, all groups had an increase in the proportion of conversations with an empowered mindset. This provides HCPs with an opportunity to drive adherence to treatment. The HBM can be used as a basis for developing culturally appropriate educational programs and materials that will increase adherence to treatment on an ongoing basis [33,36,37].

In summary, findings from this study demonstrated that individuals of all racial/ethnic groups come online to talk about depression. These findings also underscore the importance of considering how race/ethnicity might impact an individual's ability to ask for or receive mental health help and the need to make resources and information available in a culturally relevant manner and in appropriate languages, where applicable.

Limitations

Limitations of this analysis include the fact that the application of the tenets of the HBM were a priori, which may limit the replicability of these findings. Additionally, only digital conversations were considered; thus, the characteristics of those who join online communities to discuss depression may not reflect the overall population of those with depression, which may limit the reliability and generalizability of these findings. There may also be biases related to the different sources of digital conversations. The fact that only people who self-identified as a specific race/ethnicity were included in the study, that it was not possible to confirm a diagnosis of depression as patients were self-identified as having depression, and that conversations about depression were captured instead of conversations among patients with depression are also sources of possible bias. Despite these limitations, this study used advanced methodology and a large sample size to explore prevalent themes expressed by people of different racial/ethnic backgrounds and how they align with the tenets of the HBM, which can be leveraged to inform culturally competent interventions to address depression.

Conclusion

AI-powered data analysis can contribute to better health care communications and patient engagement. This study demonstrated that there were considerable racial/ethnic differences in attitudes and mindsets toward depression and drivers and barriers to seeking help and treatment pre- and mid-COVID-19. These findings may help guide initiatives that proactively educate and empower caregivers, HCPs, and families with culturally sensitive information to contextualize and address depression in their communities. Future studies should aim to replicate these findings with psychometrically validated instruments and using additional theoretical frameworks (eg, other than the HBM) and other minority groups (eg, the lesbian, gay, bisexual, transgender, queer or questioning, intersex, and asexual [LGBTQIA+] community).

Acknowledgments

This study was sponsored by Janssen Scientific Affairs, LLC, which had a role in the study design, analysis, interpretation of data; the writing of the paper; and the decision to publish. All authors were involved in the conception or design of the work and the acquisition, analysis, or interpretation of the data; critically reviewed the manuscript for content; and approved the final version to be published. Medical writing support was provided by Thomas J Parkman, PhD, MBA, of Cello Health Communications/MedErgy and was funded by Janssen Scientific Affairs, LLC.

Data Availability

The data sets generated during or analyzed during the study are not publicly available but may be available from the corresponding author on reasonable request.

Conflicts of Interest

RC-P is an employee of Janssen Research & Development, LLC, and is a stockholder in Johnson & Johnson. JP is an employee of Janssen Scientific Affairs, LLC. CB, PF, and LGV are employees of CulturIntel, which received consultancy fees from Janssen Scientific Affairs, LLC, for performing the study. TF received consultancy fees from Janssen Scientific Affairs, LLC, for this study.

References

1. G.B.D. Disease Injury Incidence Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018 Nov 10;392(10159):1789-1858 [FREE Full text] [doi: 10.1016/S0140-6736(18)32279-7] [Medline: 30496104]
2. Villarroel MA, Terlizzi EP. Symptoms of depression among adults: United States, 2019. NCHS Data Brief 2020 Sep(379):1-8. [Medline: 33054920]
3. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders: DSM-5. Washington, DC: American Psychiatric Association; 2013.
4. Ettman CK, Abdalla SM, Cohen GH, Sampson L, Vivier PM, Galea S. Prevalence of depression symptoms in us adults before and during the COVID-19 pandemic. JAMA Netw Open 2020 Sep 01;3(9):e2019686 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.19686] [Medline: 32876685]
5. Alegría M, Chatterji P, Wells K, Cao Z, Chen C, Takeuchi D, et al. Disparity in depression treatment among racial and ethnic minority populations in the United States. Psychiatr Serv 2008 Nov;59(11):1264-1272 [FREE Full text] [doi: 10.1176/appi.ps.59.11.1264] [Medline: 18971402]
6. Givens JL, Houston TK, Van Voorhees BW, Ford DE, Cooper LA. Ethnicity and preferences for depression treatment. Gen Hosp Psychiatry 2007;29(3):182-191. [doi: 10.1016/j.genhosppsych.2006.11.002] [Medline: 17484934]
7. Schraufnagel TJ, Wagner AW, Miranda J, Roy-Byrne PP. Treating minority patients with depression and anxiety: what does the evidence tell us? Gen Hosp Psychiatry 2006;28(1):27-36. [doi: 10.1016/j.genhosppsych.2005.07.002] [Medline: 16377362]
8. Eghaneyan BH, Killian MO, Sanchez K, Cabassa LJ. Psychometric properties of three measures of stigma among Hispanics with depression. J Immigr Minor Health 2021 Oct;23(5):946-955. [doi: 10.1007/s10903-021-01234-3] [Medline: 34152503]
9. Nelson T, Ernst SC, Tirado C, Fisse JL, Moreno O. Psychological distress and attitudes toward seeking professional psychological services among black women: the role of past mental health treatment. J Racial Ethn Health Disparities 2022 Apr 08;9(2):527-537. [doi: 10.1007/s40615-021-00983-z] [Medline: 33559107]
10. Ojeda VD, McGuire TG. Gender and racial/ethnic differences in use of outpatient mental health and substance use services by depressed adults. Psychiatr Q 2006;77(3):211-222. [doi: 10.1007/s11126-006-9008-9] [Medline: 16927167]
11. Alegría M, McGuire T. Rethinking a universal framework in the psychiatric symptom-disorder relationship. J Health Soc Behav 2003 Sep;44(3):257-274. [Medline: 14582307]
12. Whaley AL. Cultural mistrust and mental health services for African Americans. Couns Psychol 2016 Jun 30;29(4):513-531. [doi: 10.1177/0011000001294003]
13. Misra S, Jackson VW, Chong J, Choe K, Tay C, Wong J, et al. Systematic review of cultural aspects of stigma and mental illness among racial and ethnic minority groups in the United States: implications for interventions. Am J Community Psychol 2021 Dec 03;68(3-4):486-512. [doi: 10.1002/ajcp.12516] [Medline: 33811676]
14. Echeverry J. Treatment Barriers Accessing and Accepting Professional Help. Psychological Interventions and Research with Latino Populations. Needham Heights, MA: Allyn & Bacon; 1997:94-107.
15. Garcia ME, Hinton L, Gregorich SE, Livaudais-Toman J, Kaplan C, Karliner L. Unmet mental health need among Chinese and Latino primary care patients: intersection of ethnicity, gender, and English proficiency. J Gen Intern Med 2020 Apr;35(4):1245-1251 [FREE Full text] [doi: 10.1007/s11606-019-05483-9] [Medline: 31667737]

16. Lopez V, Sanchez K, Killian MO, Eghaneyan BH. Depression screening and education: an examination of mental health literacy and stigma in a sample of Hispanic women. BMC Public Health 2018 May 22;18(1):646 [FREE Full text] [doi: 10.1186/s12889-018-5516-4] [Medline: 29788998]
17. Kuramoto-Crawford SJ, Han B, McKeon RT. Self-reported reasons for not receiving mental health treatment in adults with serious suicidal thoughts. J Clin Psychiatry 2017 Jun;78(6):e631-e637. [doi: 10.4088/JCP.16m10989] [Medline: 28406268]
18. Yellow Horse AJ, Vargas ED. Legal status, worries about deportation, and depression among Asian immigrants. J Immigr Minor Health 2021 Jul 29:Epub ahead of print. [doi: 10.1007/s10903-021-01252-1] [Medline: 34324125]
19. Castonguay J, Filer CR, Pitts MJ. Seeking help for depression: applying the health belief model to illness narratives. South Commun J 2016 Jun 07;81(5):289-303. [doi: 10.1080/1041794x.2016.1165729]
20. Glanz K, Rimer BK, Viswanath K. The scope of health behavior and health education. In: Health Behavior and Health Education: Theory, Research, and Practice, 3rd Ed. San Francisco: Jossey-Bass; 2002:3-21.
21. Janz NK, Champion VL, Strecher VJ. The health belief model. In: Health Behavior and Health Education: Theory, Research, and Practice, 4th Ed. San Francisco: Jossey-Bass; 2008:45-65.
22. Babu NV, Kanaga EGM. Sentiment analysis in social media data for depression detection using artificial intelligence: a review. SN Comput Sci 2022;3(1):74 [FREE Full text] [doi: 10.1007/s42979-021-00958-1] [Medline: 34816124]
23. Lossio-Ventura JA, Lee AY, Hancock JT, Linos N, Linos E. Identifying silver linings during the pandemic through natural language processing. Front Psychol 2021;12:712111 [FREE Full text] [doi: 10.3389/fpsyg.2021.712111] [Medline: 34539512]
24. Oyebode O, Ndulue C, Adib A, Mulchandani D, Suruliraj B, Orji FA, et al. Health, psychosocial, and social issues emanating from the COVID-19 pandemic based on social media comments: text mining and thematic analysis approach. JMIR Med Inform 2021 Apr 06;9(4):e22734 [FREE Full text] [doi: 10.2196/22734] [Medline: 33684052]
25. Yu JS, Xue AY, Redei EE, Bagheri N. A support vector machine model provides an accurate transcript-level-based diagnostic for major depressive disorder. Transl Psychiatry 2016 Oct 25;6(10):e931 [FREE Full text] [doi: 10.1038/tp.2016.198] [Medline: 27779627]
26. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
27. Lilly FRW, Jun H, Alvarez P, Owens J, Malloy L, Bruce-Bojo M, et al. Pathways from health beliefs to treatment utilization for severe depression. Brain Behav 2020 Dec;10(12):e01873 [FREE Full text] [doi: 10.1002/brb3.1873] [Medline: 33026186]
28. Chen E, Lerman K, Ferrara E. Tracking social media discourse about the COVID-19 pandemic: development of a public coronavirus Twitter data set. JMIR Public Health Surveill 2020 May 29;6(2):e19273 [FREE Full text] [doi: 10.2196/19273] [Medline: 32427106]
29. Hou F, Bi F, Jiao R, Luo D, Song K. Gender differences of depression and anxiety among social media users during the COVID-19 outbreak in China:a cross-sectional study. BMC Public Health 2020 Nov 04;20(1):1648 [FREE Full text] [doi: 10.1186/s12889-020-09738-7] [Medline: 33148202]
30. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr 23;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
31. Falcone T, Dagar A, Castilla-Puentes RC, Anand A, Brethenoux C, Valleta LG, et al. Digital conversations about suicide among teenagers and adults with epilepsy: a big-data, machine learning analysis. Epilepsia 2020 May;61(5):951-958 [FREE Full text] [doi: 10.1111/epi.16507] [Medline: 32383797]
32. Roy A, Nikolitch K, McGinn R, Jinah S, Klement W, Kaminsky ZA. A machine learning approach predicts future risk to suicidal ideation from social media data. NPJ Digit Med 2020 May 26;3(1):78 [FREE Full text] [doi: 10.1038/s41746-020-0287-6] [Medline: 32509975]
33. Umeukeje EM, Wild MG, Maripuri S, Davidson T, Rutherford M, Abdel-Kader K, et al. Black Americans' perspectives of barriers and facilitators of community screening for kidney disease. Clin J Am Soc Nephrol 2018 Apr 06;13(4):551-559 [FREE Full text] [doi: 10.2215/CJN.07580717] [Medline: 29545381]
34. Kirkpatrick L, Watson MR, Serrano A, Campoli M, Kaltman SI, Talisman N, et al. Primary care providers' perspectives on prescribing antidepressant medication to Latino immigrant patients: a preliminary study. J Nerv Ment Dis 2020 Mar;208(3):238-244. [doi: 10.1097/NMD.0000000000001085] [Medline: 31904669]
35. Knaak S, Mantler E, Szeto A. Mental illness-related stigma in healthcare: barriers to access and care and evidence-based solutions. Healthc Manage Forum 2017 Mar;30(2):111-116 [FREE Full text] [doi: 10.1177/0840470416679413] [Medline: 28929889]
36. James DCS, Pobee JW, Oxidine D, Brown L, Joshi G. Using the health belief model to develop culturally appropriate weight-management materials for African-American women. J Acad Nutr Diet 2012 May;112(5):664-670. [doi: 10.1016/j.jand.2012.02.003] [Medline: 22709771]
37. Scarinci IC, Bandura L, Hidalgo B, Cherrington A. Development of a theory-based (PEN-3 and Health Belief Model), culturally relevant intervention on cervical cancer prevention among Latina immigrants using intervention mapping. Health Promot Pract 2012 Jan;13(1):29-40 [FREE Full text] [doi: 10.1177/1524839910366416] [Medline: 21422254]

Abbreviations

AI: artificial intelligence
AS: Asian Americans
BA: Black Americans
HBM: health belief model
HCP: health care professional
NHW: non-Hispanic Whites
NLP: natural language processing

Edited by A Mavragani; submitted 16.09.21; peer-reviewed by G Golden, K Lerman; comments to author 04.11.21; revised version received 25.01.22; accepted 22.02.22; published 20.06.22.

Please cite as:
Castilla-Puentes R, Pesa J, Brethenoux C, Furey P, Gil Valletta L, Falcone T
Applying the Health Belief Model to Characterize Racial/Ethnic Differences in Digital Conversations Related to Depression Pre- and Mid-COVID-19: Descriptive Analysis
JMIR Form Res 2022;6(6):e33637
URL: https://formative.jmir.org/2022/6/e33637
doi:10.2196/33637
PMID:35275834

Viewpoint

Antifragile Behavior Change Through Digital Health Behavior Change Interventions

Benjamin T Kaveladze[1], MA; Sean D Young[2,3], PhD; Stephen M Schueller[1,3], PhD

[1]Department of Psychological Science, University of California, Irvine, CA, United States

[2]Department of Emergency Medicine, University of California, Irvine, CA, United States

[3]Department of Informatics, University of California, Irvine, CA, United States

Corresponding Author:
Benjamin T Kaveladze, MA
Department of Psychological Science
University of California
4201 Social & Behavioral Sciences Gateway
Irvine, CA, 92697-7085
United States
Phone: 1 310 984 9318
Email: bkavelad@uci.edu

Abstract

Digital health behavior change interventions (DHBCIs) offer users accessible support, yet their promise to improve health behaviors at scale has not been met. One reason for this unmet potential may be a failure to offer users support that is tailored to their personal characteristics and goals. We apply the concept of antifragility to propose how DHBCIs could be better designed to support diverse users' behavior change journeys. We first define antifragility as a feature of an individual's relationship to a particular challenge such that if one is antifragile to a challenge, one is well positioned to benefit from facing that challenge. Second, we introduce antifragile behavior change to describe behavior change processes that leverage person-specific antifragilities to maximize benefits and minimize risk in the behavior change process. While most existing behavior change models focus on improving one's motivation and ability to face challenges, antifragile behavior change complements these models by helping to select challenges that are most likely to produce desired outcomes. Next, we propose three principles by which DHBCIs can help users to develop antifragile behavior change strategies: providing personalized guidance, embracing variance and exploration in choosing behaviors, and prioritizing user agency. Finally, we offer an example of how a DHBCI could be designed to support antifragile behavior change.

(JMIR Form Res 2022;6(6):e32571) doi:10.2196/32571

KEYWORDS

digital health behavior change interventions; behavior change; digital health; self-management; antifragile

The Need for More Effective Digital Health Behavior Change Interventions

Improving health behaviors such as eating, exercise, and sleep can profoundly impact one's life. Nonetheless, behavior change is a notoriously difficult and deeply personal process. To support health behavior change, a variety of digital tools and services, called digital health behavior change interventions (DHBCIs), have been developed. Millions of people around the world already use DHBCIs via apps, websites, and wearables, so making these tools maximally beneficial and minimally risky for users is an important goal [1].

Despite enthusiasm about DHBCIs' potential to improve people's health at scale, randomized controlled trials of DHBCIs have not provided strong evidence of improvements in health behaviors or outcomes [2]. Furthermore, a meta-analysis found that physical activity–based DHBCIs were helpful for users with high socioeconomic status (SES) but not for users with low SES, suggesting that more attention is needed to ensure that DHBCIs can support diverse populations [3]. Finally, most users do not engage with DHBCIs sufficiently to achieve intended outcomes [4,5]. In short, DHBCIs are less effective, less engaging, and less accessible than hoped. However, researchers have argued that greater integration of the behavior change theory, aligning with the affordances and limitations of mobile platforms, could improve DHBCIs' effectiveness [2,6,7].

One reason for DHBCIs' unmet potential may be their inability to offer dynamic support that accounts for users' personal strengths, weaknesses, and goals. We propose that future DHBCIs will be more successful if they can provide users with information and motivation that is contoured to their needs, helping them make the most of their opportunities to gain from while avoiding barriers and excessive risks. Here, we present the term "antifragile behavior change" to define behavior change processes that are made more efficient by leveraging user-specific factors. Antifragility is a concept that has been applied to economic and biological systems, but its application to behavior change can help identify new ways to tailor and personalize DHBCIs. As such, we put forward principles for designing DHBCIs to support antifragile behavior change and offer an example of a DHBCI that aligns with these principles.

Antifragile Behavior Change

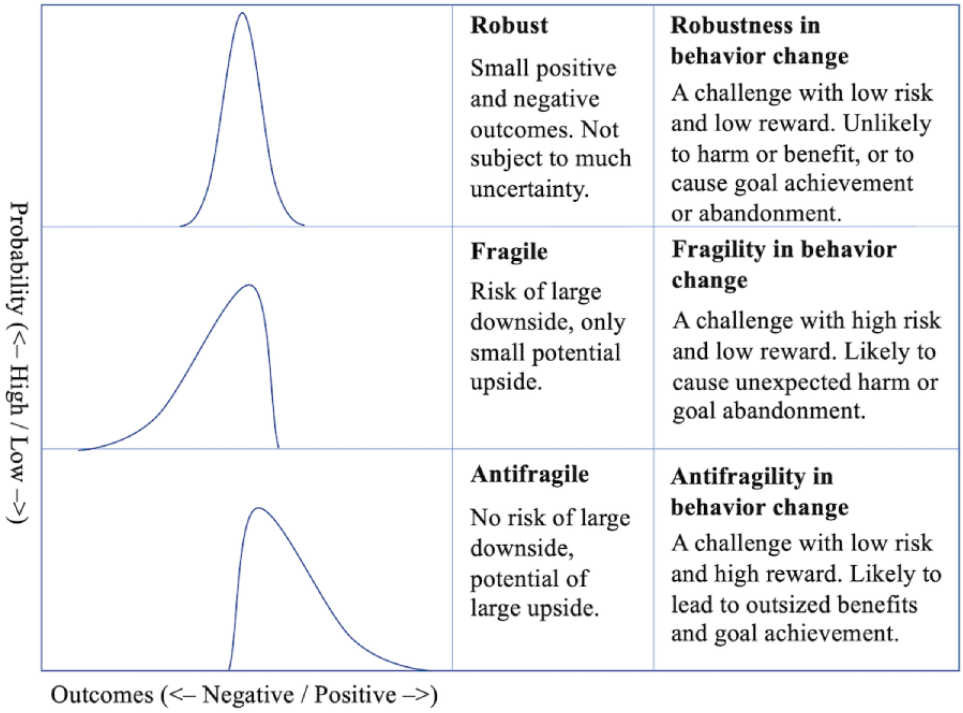
Antifragility

"Antifragile" is a term coined by Nassim Nicholas Taleb to describe systems that gain from disorder, stressors, or uncertainty (broadly, challenges) [8]. With persistent exposure to a particular challenge, things that are fragile to that challenge break, those that are robust to it stay intact, and those that are antifragile to it get better. Antifragility is not a mindset or other kind of psychological process—it only describes one's objective likelihood of gain and loss from facing a specific challenge. As a property of individuals, antifragility is a feature of the relation between a person and a particular challenge; all else equal, a student who grows and improves from criticism of their work (ie, is antifragile to the challenge of criticism) will tend to outperform a student who crumbles under (ie, is fragile to) criticism as well as a student who ignores (ie, is robust to) criticism. Nonetheless, a student who is antifragile to criticism may also be fragile to other challenges, such as focusing during long classes.

Antifragility in Behavior Change

Behavior change involves persistent and varied challenges, such as difficult workouts, diets, and competing time commitments. One's level of antifragility, robustness, or fragility to a challenge determines the outcome of their exposure to it. If one is antifragile to a challenge, one is likely to derive outsized benefits from exposure to that challenge compared to their risk of harm. Inversely, if one is fragile to a challenge, exposure risks significant downsides and only affords minor possible benefits (Figure 1). Therefore, effectively choosing the challenges one faces is critical to successful behavior change: spending too much effort on challenges to which one is fragile is inefficient and potentially counterproductive, while not spending enough effort on challenges to which one is antifragile forgoes opportunity for gain.

Figure 1. Mapping probability distributions to a behavior change context, positive (favorable) outcomes are viewed as progress toward successful behavior change or other benefits and negative (unfavorable) outcomes as movement toward giving up on one's behavior change goal or other unintended consequences. In this model, each challenge has a unique probability distribution based on individual and contextual factors. As an individual continues exposure to a challenge, their outcomes can be seen as repeated samples from the challenge's distribution.



Of course, it can be difficult to determine if one is fragile, robust, or antifragile to a given challenge; this uncertainty is inherent to the behavior change process. Exposure to a challenge can impact several domains of one's life over varying timespans, so the benefits and risks of a challenge may be hard to predict and detect. Moreover, one's relation to challenges is dynamic, changing as one progresses through the behavior change process and alongside various other life factors. Another source of this uncertainty is personal and situational context; one cannot be labeled as fragile or antifragile to a challenge without an understanding of the context in which one faces that challenge. For example, the potential outcomes of a 30-minute walk outside

differ depending on factors such as one's walking ability and their neighborhood's safety. Addressing the meta-challenge of choosing which challenges to pursue, how much, and how often, is key for successful behavior change.

Defining Antifragile Behavior Change

We define antifragile behavior change as a behavior change strategy wherein one faces challenges to which they are antifragile and avoids those to which they are fragile. We claim that such a strategy should be maximally efficient in aiding progress toward one's behavior change goal while avoiding setbacks and other negative outcomes. Although the concept is intuitive (put most simply, "do what works for you, don't do what doesn't"), creating and maintaining an antifragile behavior change strategy requires detailed self-knowledge, careful monitoring of outcomes, and possibly expert guidance to help identify one's opportunities, risks, and blind spots.

Relation to Other Behavior Change Strategies

Antifragile behavior change provides a unique perspective that complements other models of health behavior change. While existing behavior change models generally focus on improving one's motivation and ability to face challenges, antifragile behavior change asserts that these efforts are only useful if those challenges lead to desired outcomes. For example, models such as self-efficacy, mental contrasting, self-control, risk aversion, and a growth mindset are concerned with one's beliefs and attitudes toward challenges [9-11]. Similarly, the Fogg Behavior Model (FBM) identifies how contextual factors might trigger, facilitate, or discourage facing a challenge [12]. Other models such as the START (Specificity, Timing, Acquisition, Rewards and feedback, and Tools) model aim to help individuals set actionable goals [13]. However, all of these models take for granted that the challenges they help one to pursue will efficiently lead to desired outcomes. Antifragile behavior change fills a gap in these models by providing a framework that informs which challenges to pursue and which ones to avoid.

This is not to say that attitudes and supports for behavior change are not important. First, these models are valuable within the context of antifragile behavior change because they shape the nature of a challenge, and, as such, the costs and benefits that the challenge carries. For example, approaching the challenge of swimming laps with a growth mindset, an attitude of playfulness, and a swimming buddy changes the experience and impact of that swim. Moreover, no matter how beneficial a challenge might be for someone, one must be willing to engage with it to reap those benefits (yet, it is also true that one can be antifragile to a challenge one finds odious and fragile to one they find irresistible). In sum, improving motivation and facilitating behaviors can help an individual effectively engage with challenges, which is necessary but not sufficient for improving outcomes [4].

Antifragile Behavior Change in Practice

As an example of antifragile behavior change, a boxer training for a fight may choose to focus their efforts on sparring and meditation because they know that they are antifragile to those challenges, while also training in isolation with a strict diet because they know that they are fragile to the challenges of resisting distraction and temptation. Meanwhile, their extraverted opponent might choose to train around friends and family because they perform better with their support. Both boxers would almost certainly benefit from some training in all relevant domains, but the amount of time and effort that each boxer dedicates to each challenge should depend on their relations to those challenges.

Importantly, both boxers' strategies would have been shaped by years of working with trainers to try out different challenges and find the ones that work for them. A trainer's role is to identify their client's fragilities and antifragilities, use that information to create a strategy that will maximize gains and minimize losses, and then provide structure and motivation to support the execution of the strategy. Next, we propose how a DHBCI can play the role of a trainer for users pursuing behavior change.

Principles for Designing DHBCIs to Support Antifragile Behavior Change

Overview

DHBCIs designed to support antifragile behavior change aim to make users' behavior change journeys as productive as possible while avoiding serious downsides. We propose three principles for designing DHBCIs to support antifragile behavior change. First, to help users make the most of their antifragilities and avoid risk from their fragilities, DHBCIs should provide personalized guidance in the process of creating and implementing behavior change strategies. Second, to maximize opportunities to identify antifragilities, DHCBIs should encourage users to explore varied challenges. Third, to avoid unintended downsides, DHBCIs should prioritize user agency in decision-making by avoiding design choices that manipulate user choice and by explaining recommendations. We claim that these principles are flexible enough to be applied to a wide range of DHBCIs.

Providing Personalized Guidance

DHBCIs designed to support antifragile behavior change should guide users through their unique behavior change journeys. This guidance should draw from evidence on effective behavior change approaches and be responsive to users' ever-changing situations. To begin, DHBCIs should help users specify reasonable behavior change goals, drawing from evidence on effective goal setting and data on behavior change outcomes [14]. Next, they should assist users in selecting challenges to which they are likely to be antifragile and creating tailored behavior change plans based on those antifragilities. Finally, DHBCIs should support users in rigorously implementing their behavior change strategies, monitoring their progress, and making adjustments as necessary. This aligns closely with goal-oriented medical care, in which a patient and provider collaborate to provide care that matches a patient's needs [15].

In each of these stages, DHBCIs can offer users support through information (eg, expert tips and suggested activities), structure (eg, setting a schedule and tracking their progress), and motivation (eg, personalized reminders and a compelling user experience) [16,17]. This guidance can be provided by human

coaches, virtual conversational agents, or digital features such as text, images, or videos. Guidance should draw from scientific evidence on effective behavior change strategies and data on behavior change outcomes from similar others (if available) [16,17] and also adapt to personal factors.

Embracing Variance to Identify Antifragile Opportunities

As noted, the meta-challenge of choosing which challenges to pursue and to what extent is central in antifragile behavior change. Addressing this meta-challenge requires an ongoing and dynamic information gathering process. To approximate one's antifragilities, one can draw from evidence on behavior change and guidance from domain experts, but one should also personally engage in trial and error. This exploration is useful in the context of behavior change insofar as it can help one identify antifragilities (what Taleb calls "convex tinkering") [8]. This process is also related to the exploration-exploitation dilemma, in which one must choose between a known and unknown outcome [18]. Once an antifragility is identified, individuals can shift to an exploitation strategy to maximize benefits in that area.

In light of this, DHBCIs should encourage users to continuously explore potential areas of antifragility, as long as such exploration is safe and potentially fruitful. Many current DHBCIs operate through repetition; for example, a DHBCI might send a user the same daily reminder or encourage a few features for repeat use. Instead, DHBCIs should promote exploration by leveraging variety in their design; this might involve offering a range of reminders, features, and interactions to help users find the approaches that work best for them. In addition, DHBCIs should support users in monitoring the impacts of their exploration by prompting user reflection and supporting data input. Importantly, exploration carries risks that should be weighed against its potential benefits; as we discuss below, a DHBCI should not introduce variation in a way that reduces user control.

Prioritizing User Agency

Every behavior change journey involves a distinct set of challenges and facing any challenge may have a complex range of impacts on one's life. Thus, decision-making in behavior change requires a holistic and detailed understanding of the person engaged in behavior change. Because one can likely predict and monitor outcomes from one's own behavior change efforts better than an outside observer, one should have control over the decision-making process in their behavior change journey—this, again, aligns with goal-oriented medical care, which prioritizes patient decision-making over physician judgement [15]. This is true whether the external observer is a human or an automated system; while smartphone usage data can provide granular insights into a user's behavior and mental state, it is far from providing a holistic understanding of a user's life [19,20]. As such, we conclude that DHBCIs must be careful to provide support without undermining user agency.

Many DHBCIs employ nontransparent nudges, which attempt to relieve users of the psychological friction of making difficult choices by nudging them toward desired behaviors without their awareness [21]. In reducing users' agency without sufficient information to reliably predict outcomes, these nontransparent nudges put users at risk of unintended downsides. The adverse spillover and second-order effects of these nudges into other domains of users' lives are extremely difficult to predict and potentially costly [8,22]; a DHBCI designed to optimize one's fitness will not simultaneously optimize one's overall well-being or one's roles as a friend and parent. Thus, efforts to shift user behavior without continuous user consent are inconsistent with antifragile behavior change.

In addition to avoiding nontransparent nudging, DHBCIs also need to encourage users to interact with their content critically. Specifically, DHBCIs should provide explanations to justify and contextualize their recommendations. For example, if a smartwatch running app suggests that a user take a rest day, it should explain why it is doing so and indicate its confidence in that suggestion (eg, "for most people with your level of experience, taking a rest day after training as hard as you did today is moderately helpful for building long-term endurance and avoiding injuries"). Without such explanations, a user might overestimate the accuracy or impact of a recommendation that is ultimately not well suited to their needs. In addition to prioritizing user agency in decision-making, explaining recommendations can increase DHBCIs' effectiveness and trustworthiness [23].

As in the Greek myth of Procrustes, who, in an attempt to give each of his guests a perfect night's sleep, stretches them out or amputates their legs to make them exactly fit his bed's length, a DHBCI that attempts to optimize behavior change from the outside can misjudge users' needs and cause unintended consequences [8]. As DHBCIs become more ubiquitous and engaging in the coming years, eliminating the risk of iatrogenic effects will be increasingly important. DHBCIs designed for antifragility should avoid using nontransparent nudges and offering recommendations without explaining them. Instead, they should prioritize user agency by providing transparent support that reflects the best available evidence, while also explaining the logic, strengths, and limitations of that support. Such guidance can empower users to make informed decisions within the dynamic contexts of their lives.

Based on the design principles we proposed—personalizing guidance, embracing variance, and prioritizing user agency—we offer a brief conceptualization of how well a DHBCI supports antifragile behavior change in Textbox 1.

Textbox 1. Considerations to determine how well a digital health behavior change intervention aligns with our principles of design to support antifragile behavior change.

Does the DHBCI…

- Principle: personalize guidance
  - Help users consider possible costs and benefits of challenges?
  - Help users monitor the impacts of their choices over time?
  - Allow users to personalize their behavior change strategies?
  - Change based on new information or feedback from users?
- Principle: embrace variation and exploration
  - Encourage users to try new challenges or interactions?
  - Introduce a variety of features over time to support exploration?
  - Prompt user reflection on potentially unhelpful patterns?
- Principle: prioritize user agency
  - Avoid nudges that could influence user behavior without consent?
  - Appropriately describe its limitations and uncertainty?
  - Provide users clear explanations for its recommendations?

Alignment With and Differences From Other Perspectives on DHBCIs

App Behavior Change Scale

Our design principles are consistent with some of the recommendations for DHBCI design outlined in the App Behavior Change Scale (ABACUS), a tool used to determine the behavior change potential of apps [24]. Specifically, our model aligns with the ABACUS's recommendations that apps should allow users to "customize and personalize features," provide "information about the consequences of continuing and discontinuing behavior," help with "goal setting," and offer "opportunity to plan for barriers" [24]. Antifragile behavior change explains that this personalization is valuable because the costs and benefits derived from challenges are person- and context-specific.

FBM

The FBM, which is based on the observation that motivation, ability, and a prompt are necessary for a desired behavior to take place, has been influential in behavior change and user experience design more generally [12]. Some interpretations and applications of the FBM attempt to support behavior change by promoting nontransparent nudging or other ways of influencing behavior without consent [21]. Our design principle of prioritizing user agency is opposed to such approaches. Nonetheless, the FBM also provides strategies for improving user motivation and simplifying tasks that could be useful for behavior change. We argue that integrating strategies from the FBM, such as increasing user ability to complete behaviors, considering timing and triggers, and addressing core motivators, can be helpful as long as those design choices are employed transparently and do not infringe on user agency [12].

Efficiency Model of Support

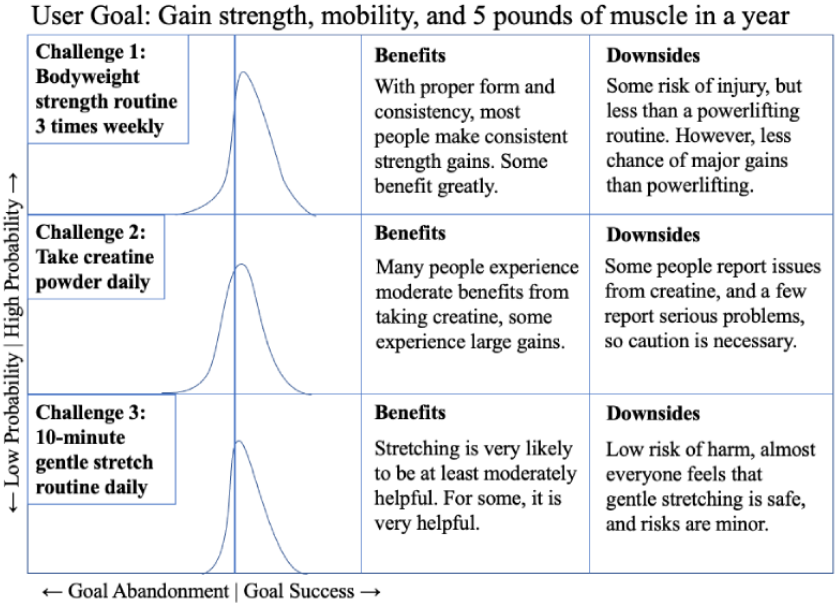
Our approach parallels Schueller, Tomasino, and Mohr's Efficiency Model of Support [25], which considers ways to improve the efficiency of human support in behavioral intervention technologies (efficiency is defined as the ratio of benefit accrued to resources devoted to providing that support). Antifragile behavior change concerns the efficiency of the behavior change strategy as a whole: an efficient strategy maximizes progress toward one's goal while minimizing the chance of abandoning the goal, as visualized in the antifragile distribution in Figure 1. In addition, consistent with our focus on person-specific strategies, the efficiency model of support emphasizes that the ideal type of support is contingent on person-level characteristics and goals [25].

Example of a DHBCI Designed to Support Antifragile Behavior Change

Here we sketch an outline for a mobile app–based DHBCI designed to support antifragile behavior change for building physical strength. Our example draws from existing approaches to technology-supported behavior change, including behavior change techniques such as social support and instruction on performing behaviors [16,17,26]. It also aligns with the conceptual and technological architecture proposed by the Behavioral Intervention Technology Model [27]. Finally, it follows McCallum et al's [28] "n-of-1 evaluation framework for behavior change applications" to inform which user variables are collected, aiming to gain information that can be aggregated and used to improve tailoring with high internal, external, and social validity [28]. Rather than introducing new features, our aim is to demonstrate how our design principles might be implemented in a DHBCI similar to existing ones by highlighting or modifying features with demonstrated real-world effectiveness.

The app begins with a planning phase, in which the user generates a behavior change strategy, followed by an execution or evaluation phase, in which the user provides data to monitor their progress and adjusts their strategy if needed. The app's user interface centers on visualizations of predicted outcomes of each challenge, in which "prior" predicted outcomes on the basis of information crowdsourced from other users are updated as a user reports their experiences, forming increasingly personalized "posterior" probability distributions (Figure 2).

Figure 2. Three challenges a user might select to incorporate in their behavior change strategy during the planning phase of our digital health behavior change intervention example. The app recommends these challenges to a user based on challenge outcome data reported by other app users with similar characteristics and goals. The initial probability distributions shown are based on outcome data from other users, but as a user reports outcomes from their efforts, their predicted outcome distribution for each challenge is updated.



To begin the planning phase, the app provides the following brief introduction:

> *In the words of C.T. Vivian, 'You are made by the struggles you choose.' To change your behavior, you have to put in effort to face challenges, such as intense workouts or maintaining a good sleep schedule. Choosing the right challenges for you will bring you closer to your goal, but facing the wrong challenges could hold you back or even cause injury. This app will help you to find and maintain the right behavior change strategy for you, but remember that you know yourself best, so you should always consider the app's recommendations in the context of what is right for you and those around you.*

Note that although this introduction explicitly discusses concepts relevant to antifragile behavior change, a DHBCI can be consistent with our design principles without doing so.

Continuing the planning phase, the app asks the user for some personal context, such as their behavior change goal, goal timeline, age, and experience with physical exercise. Based on this information and data from similar people who used the app previously, it presents a series of suggested behavior change challenges that the user can select from to create their strategy. These challenges might include a regular weight lifting regimen, a stretching routine, a macronutrient goal, and refraining from drinking alcohol. Figure 2 provides an example of how these challenges might be presented to users, including distributions to help users visualize risks and opportunities from each challenge. Further, the app provides a series of support options for the user to choose from; for example, motivational nudges and personalized recommendations from the app, instruction on challenges, or access to a peer support community to exchange advice and encouragement. These support options can be further customized, allowing the user to choose the app's tone (eg, kind and funny or tough and demanding) or their support community's makeup (eg, women from the age group of 18-30 years or people interested in bodybuilding). Finally, the app creates an editable schedule with all of the user's selected challenges.

During the execution or evaluation phase, the user completes challenges when scheduled and reports positive and negative impacts that they believe to have resulted from each challenge. These data are used to generate personalized suggestions for adjusting one's behavior change strategy and to update their predicted outcomes. For example, if a user consistently experiences more downsides than upsides after weight training, the app might recommend power yoga to achieve the same goal from a different approach. Throughout this phase, the app also encourages the user to continue exploring new challenges to discover new areas of antifragility. Finally, the user's outcome data are aggregated with other users' data to contribute to the app's knowledge base of behavior change strategies. Because it gains aggregate information from many users' successes and failures, the app itself can be seen as antifragile to its users' efforts.

Figure 2 shows the three challenges a user might select to incorporate in their behavior change strategy during the planning phase of our DHBCI example. The app recommends these

challenges to a user on the basis of challenge outcome data reported by other app users with similar characteristics and goals. The initial probability distributions shown are based on outcome data from other users, but as a user reports outcomes from their efforts, their predicted outcome distribution for each challenge is updated.

Future Directions

Our paper presents antifragile behavior change as a theoretical concept that can be useful for DHBCI design. Future work can test this concept and potentially validate and extend it. First, advancing the understanding of antifragile behavior change will require measures to assess whether DHBCIs conform to the principles we proposed. Textbox 1 provides some conceptual considerations of how well a digital intervention aligns with antifragile behavior change, which might be adapted into a formal measure or assessment. Once a measure is created and validated, it opens up the possibility of addressing various research questions; for example, how well do different commercially available DHBCIs use antifragile behavior change principles, and do such interventions result in greater impact or engagement? Second, both researchers and developers could integrate antifragile behavior change principles into their designs. For researchers, this would enable tests of antifragile behavior change's effectiveness: comparing DHBCIs and DHBCI features that align with the principles of antifragile behavior change to those that do not in terms of behavior change, health, and engagement outcomes. For developers, this represents another way to incorporate conceptually grounded behavior change techniques into their products. Third, research is needed to test our design principles across different DHBCIs and populations, examining the externalities of antifragile behavior change and trade-offs between strategies. For example, removing design elements that could threaten user agency might make it harder to tailor content, thereby decreasing some users' willingness to engage with an app's recommendations. These directions could help extend our suggestions into research and applications for the field.

Conclusions

DHBCIs hold great potential to help people improve their health-related behaviors and overall well-being, but their promise has been held back by a failure to provide support that fits the contours of diverse users' lives. Viewing DHBCIs through the lens of antifragility reveals both opportunities for advancement and reasons for caution. We presented three principles for designing DHBCIs to support antifragile behavior change. First, DHBCIs should provide personalized guidance to fit user-specific fragilities and antifragilities throughout the behavior change process. Second, DHBCIs should encourage users to explore varied challenges to discover new areas of antifragility. Third, DHBCIs should not limit user control in an attempt to make behavior change easier, as such efforts to optimize users' lives from the top down can inadvertently cause harm. We claim that integrating these principles into DHBCI design will improve these tools' abilities to support users across diverse populations. We hope that this paper will spark interest in reconsidering DHBCIs from the perspective of antifragile behavior change, both among DHBCI developers and behavior change researchers.

Conflicts of Interest

SDY has received royalties from HarperCollins Publishers and Penguin Books for his book, *Stick with It: A Scientifically Proven Process for Changing Your Life-for Good*. He has received compensation from the company ElevateU for consulting, and equity for advising digital health startups. SMS serves on the Scientific Advisory Board for Headspace, for which he receives compensation. He has received consulting payments from Otsuka Pharmaceuticals and K Health (Trusst).

References

1. McKay FH, Wright A, Shill J, Stephens H, Uccellini M. Using Health and Well-Being Apps for Behavior Change: A Systematic Search and Rating of Apps. JMIR Mhealth Uhealth 2019 Jul 04;7(7):e11926 [FREE Full text] [doi: 10.2196/11926] [Medline: 31274112]
2. Milne-Ives M, Lam C, De Cock C, Van Velthoven MH, Meinert E. Mobile Apps for Health Behavior Change in Physical Activity, Diet, Drug and Alcohol Use, and Mental Health: Systematic Review. JMIR Mhealth Uhealth 2020 Mar 18;8(3):e17046 [FREE Full text] [doi: 10.2196/17046] [Medline: 32186518]
3. Western MJ, Armstrong MEG, Islam I, Morgan K, Jones UF, Kelson MJ. The effectiveness of digital interventions for increasing physical activity in individuals of low socioeconomic status: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2021 Nov 09;18(1):148 [FREE Full text] [doi: 10.1186/s12966-021-01218-4] [Medline: 34753490]
4. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and Promoting Effective Engagement With Digital Behavior Change Interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
5. Baumel A, Muench F, Edan S, Kane JM. Objective User Engagement With Mental Health Apps: Systematic Search and Panel-Based Usage Analysis. J Med Internet Res 2019 Sep 25;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
6. Young SD. Stick with It: A Scientifically Proven Process for Changing Your Life-for Good. New York, NY: HarperCollins; 20, 2017.

7. Young SD. The Adaptive Behavioral Components (ABC) Model for planning longitudinal behavioral technology-based health interventions: a theoretical framework. Journal of Medical Internet Research 2020 Jun 26;22(6):e15563 [FREE Full text] [doi: 10.2196/15563] [Medline: 32589152]
8. Taleb N. Antifragile: Things That Gain from Disorder. New York, NY: Random House; 2012.
9. Bandura A, Freeman WH, Lightsey R. Self-Efficacy: The Exercise of Control. J Cogn Psychother 1999 Jan 01;13(2):158-166. [doi: 10.1891/0889-8391.13.2.158]
10. Yeager DS, Dweck CS. Mindsets That Promote Resilience: When Students Believe That Personal Characteristics Can Be Developed. Educ Psychol 2012 Oct;47(4):302-314. [doi: 10.1080/00461520.2012.722805]
11. Crum AJ, Langer EJ. Mind-set matters: exercise and the placebo effect. Psychol Sci 2007 Feb;18(2):165-171. [doi: 10.1111/j.1467-9280.2007.01867.x] [Medline: 17425538]
12. Fogg B. A behavior model for persuasive design. 2009 Presented at: Persuasive 2009: Persuasive 2009; 4th International Conference on Persuasive Technology; April 26-29, 2009; Claremont, CA. [doi: 10.1145/1541948.1541999]
13. Pearson ES. Goal setting as a health behavior change strategy in overweight and obese adults: a systematic literature review examining intervention components. Patient Educ Couns 2012 Apr;87(1):32-42. [doi: 10.1016/j.pec.2011.07.018] [Medline: 21852063]
14. Latham G. Goal Setting: A five-step approach to behavior change. In: Organizational Collaboration: Themes and issues. London: Routledge; 2011.
15. Mold JW, Blake GH, Becker LA. Goal-oriented medical care. Fam Med 1991 Jan;23(1):46-51. [Medline: 2001782]
16. Embry DD, Biglan A. Evidence-based kernels: fundamental units of behavioral influence. Clin Child Fam Psychol Rev 2008 Sep;11(3):75-113 [FREE Full text] [doi: 10.1007/s10567-008-0036-x] [Medline: 18712600]
17. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
18. Sutton R, Barto A. Reinforcement Learning, Second Edition. Cambridge, MA: The MIT Press; 2018.
19. Mohr DC, Zhang M, Schueller SM. Personal Sensing: Understanding Mental Health Using Ubiquitous Sensors and Machine Learning. Annu Rev Clin Psychol 2017 May 08;13:23-47 [FREE Full text] [doi: 10.1146/annurev-clinpsy-032816-044949] [Medline: 28375728]
20. Chancellor S, De Choudhury M. Methods in predictive techniques for mental health status on social media: a critical review. NPJ Digit Med 2020;3:43 [FREE Full text] [doi: 10.1038/s41746-020-0233-7] [Medline: 32219184]
21. Caraban A, Karapanos E, Gonçalves D, Campos P. 23 Ways to Nudge: A Review of Technology-Mediated Nudging in Human-Computer Interaction. 2019 Presented at: CHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow. [doi: 10.1145/3290605.3300733]
22. Nafziger J. Spillover effects of nudges. Econ Lett 2020 May;190:109086. [doi: 10.1016/j.econlet.2020.109086]
23. Zhang Y, Chen X. Explainable Recommendation: A Survey and New Perspectives. FNT in Information Retrieval 2020;14(1):1-101. [doi: 10.1561/1500000066]
24. McKay FH, Slykerman S, Dunn M. The App Behavior Change Scale: Creation of a Scale to Assess the Potential of Apps to Promote Behavior Change. JMIR Mhealth Uhealth 2019 Jan 25;7(1):e11130 [FREE Full text] [doi: 10.2196/11130] [Medline: 30681967]
25. Schueller SM, Tomasino KN, Mohr DC. Integrating human support into behavioral intervention technologies: The efficiency model of support. Clin Psychol 2017 Mar;24(1):27-45. [doi: 10.1037/h0101740]
26. Yang C, Maher JP, Conroy DE. Implementation of behavior change techniques in mobile applications for physical activity. Am J Prev Med 2015 Apr;48(4):452-455. [doi: 10.1016/j.amepre.2014.10.010] [Medline: 25576494]
27. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. J Med Internet Res 2014 Jun 05;16(6):e146 [FREE Full text] [doi: 10.2196/jmir.3077] [Medline: 24905070]
28. McCallum C, Rooksby J, Asadzadeh P, Gray C, Chalmers M. An N-of-1 Evaluation Framework for Behaviour Change Applications. 2019 Presented at: CHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow. [doi: 10.1145/3290607.3312923]

Abbreviations

ABACUS: App Behavior Change Scale
DHBCI: digital health behavior change intervention
FBM: Fogg Behavior Model
SES: socioeconomic status
START: Specificity, Timing, Acquisition, Rewards and feedback, and Tools

Edited by A Mavragani; submitted 02.08.21; peer-reviewed by J Nafziger, P Chow, B Chaudhry; comments to author 27.10.21; revised version received 01.03.22; accepted 08.05.22; published 03.06.22.

Please cite as:
Kaveladze BT, Young SD, Schueller SM
Antifragile Behavior Change Through Digital Health Behavior Change Interventions
JMIR Form Res 2022;6(6):e32571
URL: https://formative.jmir.org/2022/6/e32571
doi:10.2196/32571
PMID:35657665

Original Paper

COVID-19 Variant Surveillance and Social Determinants in Central Massachusetts: Development Study

Qiming Shi[1], MS; Carly Herbert[2,3], BA; Doyle V Ward[4,5], PhD; Karl Simin[6], PhD; Beth A McCormick[4,5], PhD; Richard T Ellison III[3,4], MD; Adrian H Zai[1,2], MPH, MD, PhD

[1]Center for Clinical and Translational Science, UMass Chan Medical School, Worcester, MA, United States

[2]Department of Population and Quantitative Health Sciences, UMass Chan Medical School, Worcester, MA, United States

[3]Department of Medicine, UMass Chan Medical School, Worcester, MA, United States

[4]Department of Microbiology and Physiological Systems, UMass Chan Medical School, Worcester, MA, United States

[5]Center for Microbiome Research, UMass Chan Medical School, Worcester, MA, United States

[6]Molecular, Cell, and Cancer Biology, UMass Chan Medical School, Worcester, MA, United States

Corresponding Author:
Qiming Shi, MS
Center for Clinical and Translational Science
UMass Chan Medical School
362 Plantation Street
Ambulatory Care Center, 7th Floor
Worcester, MA, 01605
United States
Phone: 1 5088568989
Email: qiming.shi@umassmed.edu

Abstract

Background: Public health scientists have used spatial tools such as web-based Geographical Information System (GIS) applications to monitor and forecast the progression of the COVID-19 pandemic and track the impact of their interventions. The ability to track SARS-CoV-2 variants and incorporate the social determinants of health with street-level granularity can facilitate the identification of local outbreaks, highlight variant-specific geospatial epidemiology, and inform effective interventions. We developed a novel dashboard, the University of Massachusetts' Graphical user interface for Geographic Information (MAGGI) variant tracking system that combines GIS, health-associated sociodemographic data, and viral genomic data to visualize the spatiotemporal incidence of SARS-CoV-2 variants with street-level resolution while safeguarding protected health information. The specificity and richness of the dashboard enhance the local understanding of variant introductions and transmissions so that appropriate public health strategies can be devised and evaluated.

Objective: We developed a web-based dashboard that simultaneously visualizes the geographic distribution of SARS-CoV-2 variants in Central Massachusetts, the social determinants of health, and vaccination data to support public health efforts to locally mitigate the impact of the COVID-19 pandemic.

Methods: MAGGI uses a server-client model–based system, enabling users to access data and visualizations via an encrypted web browser, thus securing patient health information. We integrated data from electronic medical records, SARS-CoV-2 genomic analysis, and public health resources. We developed the following functionalities into MAGGI: spatial and temporal selection capability by zip codes of interest, the detection of variant clusters, and a tool to display variant distribution by the social determinants of health. MAGGI was built on the Environmental Systems Research Institute ecosystem and is readily adaptable to monitor other infectious diseases and their variants in real-time.

Results: We created a geo-referenced database and added sociodemographic and viral genomic data to the ArcGIS dashboard that interactively displays Central Massachusetts' spatiotemporal variants distribution. Genomic epidemiologists and public health officials use MAGGI to show the occurrence of SARS-CoV-2 genomic variants at high geographic resolution and refine the display by selecting a combination of data features such as variant subtype, subject zip codes, or date of COVID-19–positive sample collection. Furthermore, they use it to scale time and space to visualize association patterns between socioeconomics, social vulnerability based on the Centers for Disease Control and Prevention's social vulnerability index, and vaccination rates.

We launched the system at the University of Massachusetts Chan Medical School to support internal research projects starting in March 2021.

Conclusions: We developed a COVID-19 variant surveillance dashboard to advance our geospatial technologies to study SARS-CoV-2 variants transmission dynamics. This real-time, GIS-based tool exemplifies how spatial informatics can support public health officials, genomics epidemiologists, infectious disease specialists, and other researchers to track and study the spread patterns of SARS-CoV-2 variants in our communities.

(JMIR Form Res 2022;6(6):e37858) doi:10.2196/37858

KEYWORDS

geographic information science; GIS; COVID-19; SARS-CoV-2; variants; surveillance; dashboard; web mapping; public health; web-based information; digital health; epidemiology

Introduction

COVID-19 has infected 219 million people resulting in over 4.5 million deaths as of October 2021 [1]. The pandemic has highlighted the role of geographic mapping technologies to provide easy-to-understand indicators such as new cases, confirmed tests, and inpatient rates using spatiotemporal visualizations [2]. For example, public agencies [3] and news media [4,5] have used these visualizations to inform the public about the spread of COVID-19 and explain why officials recommend adopting intervention measures such as mask mandate and vaccinations [6].

Health science has long leveraged Geographical Information System (GIS) spatial analysis and applications [7]. GIS offers an interactive and efficient approach to revealing meaningful patterns and associations [8] that would be otherwise difficult to visualize using traditional figures and tables. As a result, researchers are increasingly using spatial analysis to study the impacts of COVID-19 [9,10] or understand the relationship between COVID-19 cases and sociodemographic or economic data and vaccination rates in local environments [11-13].

With the constant technological evolution of GIS platforms, many geospatial online dashboards are available for monitoring COVID-19 worldwide [14,15]. The COVID-19 dashboard developed by John Hopkins University is arguably the most popular and stands out for its effectiveness at capturing new cases around the globe [16]. Meanwhile, the Centers for Disease Control and Prevention's National Healthcare Safety Network has also developed a geospatial dashboard for COVID-19 infection data analysis and prevention [17]. Most GIS systems, however, track data with coarse geographic resolution (eg, state or county). Further, few GIS systems track the spatiotemporal transition dynamics of SARS-CoV-2 variants [18].

The impact of using genomic epidemiology to monitor the COVID-19 pandemic has been profound, as it provides unprecedented detail into the appearance and global dissemination of SARS-CoV-2 variants [19-24]. In addition, phylogenetic-epidemiological analysis has enabled the reconstruction of super-spreader events [25]. Notably, these analyses have illustrated the power of combining epidemiological analytics with deep viral genome sequencing to gain fundamental insights into mutational dynamics and transmission properties [26]. Meanwhile, researchers have developed visualization tools to analyze phylogenetics spatially [27-29], albeit with a coarse geospatial resolution due to poor detail in data collection.

A detailed analysis of the demographic and social determinants of disease risk provides insights into how social characteristics impact health risks and disparities [30-32]. For many infectious diseases, populations of lower socioeconomic status tend to be associated with higher disease prevalence [33-35]. Typically, GIS dashboards fail to incorporate the socioeconomic correlates of health that could provide a deeper understanding of transmission dynamics when combined with geographic and genomic data.

This paper presents the development of the University of Massachusetts' Graphical user interface for Geographic Information (MAGGI) variant tracking system, a GIS dashboard that integrates genomics and socioeconomic data into a high-resolution geographical dashboard. Our aims for developing MAGGI were to (1) track the transition dynamics of SARS-CoV-2 variants across space and time, (2) identify geographical areas at high risk for transmission using variant cluster risk analysis, and (3) assess socioeconomic risk factors within these high-risk areas.

Methods

Ethics Approval

This study was approved by the University of Massachusetts (UMass) Chan Medical School Institution Review Board (protocol H00021561).

Data Source

Data incorporated into MAGGI originates from multiple sources, including the American Community Survey [36], Massachusetts Department of Public Health (DPH) [37], Centers for Disease Control and Prevention [38], and the UMass Chan Medical School clinical research data warehouse (RDW). The RDW combines data from the UMass Memorial electronic health record system from Epic, Allscripts, and Soarian. Next, we used the 2018 American Community Survey 5-year data to extract zip code–level social demographic data, including race, ethnicity, median family income, housing, language, poverty, and population density. We downloaded the ZIP Code Tabulation Areas (ZCTA) data from the census, and ZCTA are what we used in the MAGGI system. However, we used zip code as a proxy of ZCTA since the latter will be the same as its zip code in most cases and the term zip code is more commonly

used and known. Finally, we obtained the COVID-19 vaccination rate data from the Massachusetts DPH and the social vulnerability index from the Centers for Disease Control and Prevention [33].

Remnant positive SARS-CoV-2 test patient samples (swab and saliva) from the UMass Memorial Health Care system were collected and archived by the UMCCTS Biospecimen bank. Sample aliquots were transferred to the UMass Center for Microbiome research for SARS-CoV-2 genomic extraction and sequencing. Sequencing libraries were prepared using NEBNext ARTIC SARS-CoV-2 FS Library Prep Kits and sequenced on the Illumina NextSeq 500 platform as 75nt paired-end reads per manufacturer protocols. Sequence data were analyzed using the Cecret workflow developed at the Utah Public Health Laboratory that provides SARS-CoV-2 genomic sequence and lineage determinations [39]. The genomic results were then linked to patient data.

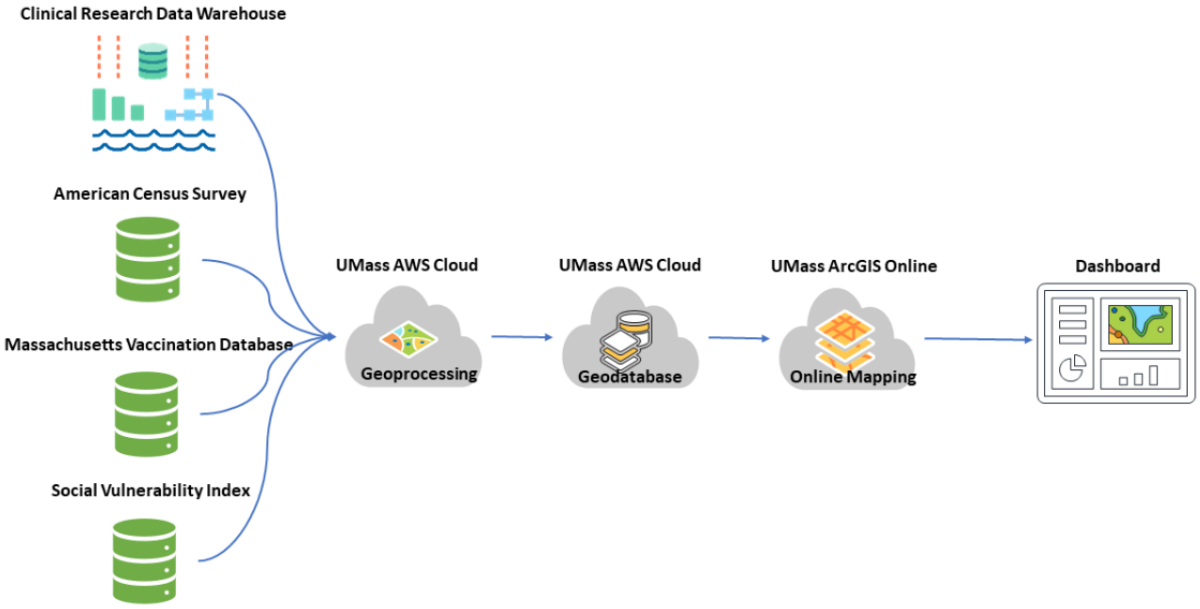
Geocoding

Geocoding is the technology used to transform physical street addresses into geographic coordinates such as latitude and longitude, enabling us to place markers on the map. After determining the SARS-CoV-2 genomic sequence and lineage, we geocoded the addresses extracted from electronic medical records using the ArcGIS Pro Geocode function on the secure UMass Amazon Web Service private cloud. All geocoding processes used a local geolocator from the Environmental Systems Research Institute (ESRI). Approximately 10%-15% of the extracted addresses contained errors or represented recent addresses that the ESRI geolocator had yet to capture; we manually corrected these errors with GPS coordinates obtained from Google Maps search results. Next, we randomly geocoded Post Office Box addresses to the area defined by the corresponding zip code. Last, we deidentified the geocoded layer to include only longitude and latitude, strain lineage, sample collection date, and variant type to further protect patient information.

Data Integration

The UMass Memorial Health Care System (UMMHC) is the primary health care provider in Central Massachusetts and has used the Epic electronic health record system [40] beginning in 2017. The Epic Clarity database is our primary source for clinical and demographic data. Viral genomic data resides in a separate database. Consequently, we created a Structured Query Language Server database to link genomic data with clinical data using a universal identifier number. This Structured Query Language database remits data to ArcGIS Pro for geocoding. The polygon geometry from zip code and census tracts was integrated into the geodatabase project. Finally, we designed the geodatabase to organize and store spatial databases, tables, and vector data sets. The geodatabase combines geocoding results, social demographics, social vulnerability, and vaccination data, and then passes the data to ArcGIS Online for web-based mapping (Figure 1).

Figure 1. Dashboard workflow and technology stack. UMass: University of Massachusetts; AWS: Amazon Web Service.

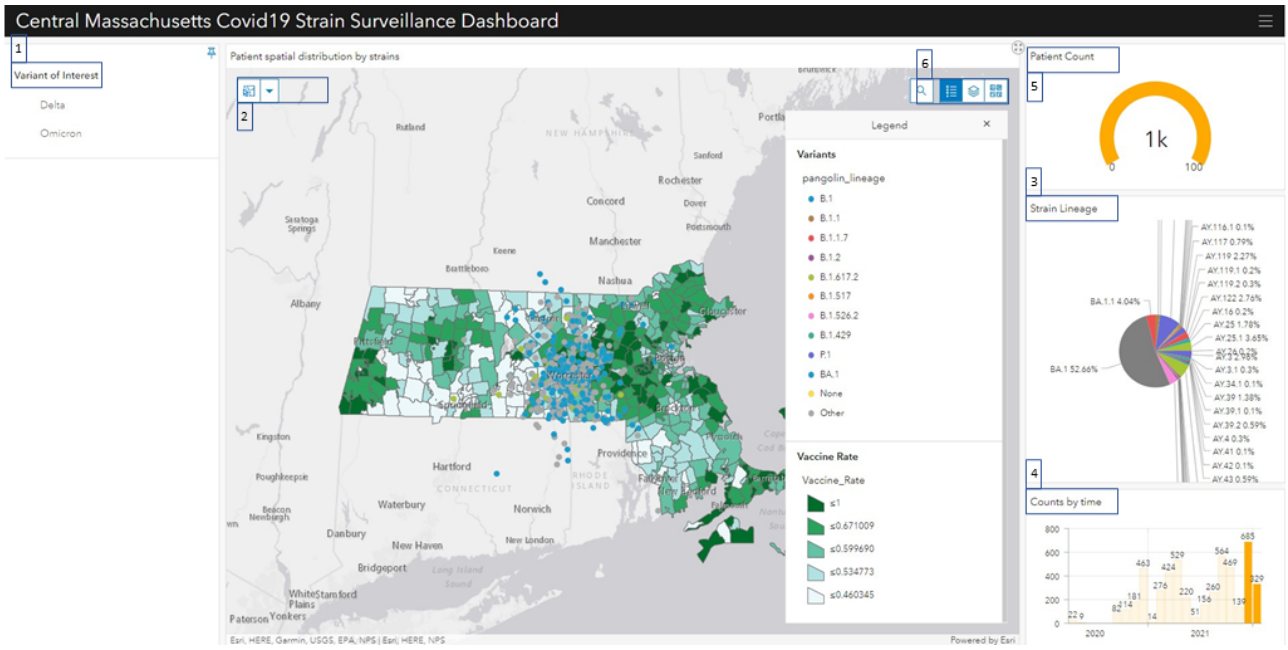


Data Visualization and Dashboard Development

ArcGIS Online, the ESRI web-based mapping software, enables users to build interactive web maps with user interface instead of JavaScript programming [41]. We created the web map to visualize spatial layers with this tool. The layers include the geocoded results of variant type, socioeconomic data, vaccination status, and social vulnerability index data. ArcGIS Dashboard, the essential component to conveying information by presenting location-based analytics using intuitive and interactive data visualization, was used to add interactive charts and functionalities based on the web map. We leveraged the dashboard's versatility to enable users to render a variety of charts and selections, which included the following features (item numbers below correspond to the boxed numbers in Figure 2).

- (1) Variants Selection Panel: We built this variant selection panel using the Category Selector function provided by the ArcGIS dashboard. This panel enables researchers to quickly filter the data to the genomic variants of interest (eg, Delta and Omicron).
- (2) Spatial Selection Tool: We leveraged the Spatial Selection method in the layer actions section provided by the ArcGIS dashboard to enable the interconnection between the genomic variants and socioeconomic layers. This function allows users to investigate patients in selected geographical units such as zip codes and census units. This tool is useful when genomic epidemiologists study and compare variant population trends through various geolocations.
- (3) Strain Lineage Pie Chart function: We created this pie chart using the ArcGIS Pie Chart function to determine the lineage distribution proportion at a given time and zip code. It helps measure the dominated variants and calculate relative risk (RR) in a specific time and space.
- (4) Time Serial Chart: We built this chart using the ArcGIS Serial Chart function to render sequenced data by time series. It displays the count of patients by month and works with other selection criteria simultaneously to study the variant transmission dynamic across space and time.
- (5) Patient Count: Using the ArcGIS Gauge function, we added the patient count gauge to track patient count based on current selection criteria.
- (6) Layer Toggle: We enabled the Basemap Switcher function in the ArcGIS dashboard. This function allows users to change the background map among the socioeconomic layers, social vulnerability index, and vaccination layer. In addition, this function helps researchers easily switch and draw potential correlations between layers.
- Cluster Detection: We used the Getis-Ord Gi* hot spot analysis [42] provided by ArcGIS Online to detect the spatial cluster of the variant.

Figure 2. The Massachusetts' Graphical user interface for Geographic Information (MAGGI) application user interface.



Data Security and Sharing

Users must log in to the ArcGIS Online application via a single sign-on with Microsoft multi-factor authentication technology [43]. Access to the dashboard is based on membership or association with the studies, ensuring that only authorized study team members will see the specific study-related dashboard upon log-in. ArcGIS Online offers several grouping categories to restrict access to appropriate users only: organization, groups within an organization, collaborators with an organization, or everyone (public) [44]. For MAGGI, we used this security protocol to set up private groups for local health department administrators to access only data from their respective jurisdictions. Thus, for example, members of the Massachusetts Fitchburg DPH group can only visualize data from the 11 towns managed by their department. In contrast, administrators from the Massachusetts DPH can access and visualize data from the entire state of Massachusetts.

Results

User Interface

An interactive dashboard provides the user interface that allows researchers to study the spatiotemporal trend of the variants as they work through the investigation process. The user interface depicting data relating to SARS-CoV-2 variants from patient infections between May and October 2001 is presented (Figure 2).

Examples of Use Cases

Use Case 1: Variant Surveillance

MAGGI enables members of public health agencies in Massachusetts and genomic epidemiologists to monitor regional viral spread patterns and the emergence of potential new variants and study the impact of new interventions, including the effectiveness of new policy implementations. For example,

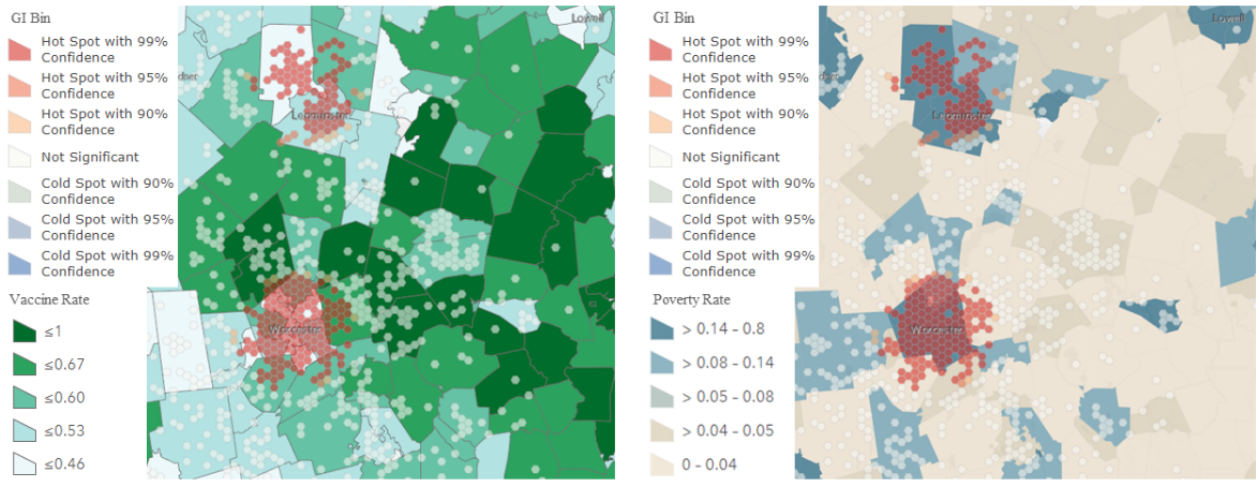
genetic epidemiologists use MAGGI to study the relationship between genetic drifts in the genome of SARS-CoV-2 and transmission rate at the local level, thereby revealing unsuspected clusters and evidence for or against suspected transmissions.

Use Case 2: Spatiotemporal Cluster Detection

The Hot Spot Analysis tool calculates the cluster risk using the Getis-Ord Gi* statistic [42]. This calculation's resulting z scores and *P* values inform us where to locate areas with high- or low-value clusters spatially. This tool functions by comparing the z scores and *P* values between any location with its neighbors. Regions labeled as "hot spots" must have a statistically significant score compared to their surrounding neighbors. Based on the cluster risk analysis (Getis-Ord Gi*) results, our first use case objective was to explore the social and vaccination determinants associated with detected clusters. Clusters were identified in the Worcester and Leominster areas served by UMMHC and co-located with areas of lower socioeconomic status and low vaccination rates (Figure 3).

Users can further investigate the dissemination of emerging variants. First, a user specifies the variant of interest and selects the zip codes from the map. The map then updates to visualize the infections meeting the criteria. The user then visualizes the distribution in the time series chart and individually inspects each interval by making the appropriate selections. Finally, the user may browse the filtered results via the map and further probe the distribution of emerging variants or clusters.

Figure 3. Hot spot analysis (Getis-Ord Gi*) layer overlapping with the social and vaccination determinants layer.



Use Case 3: Spatial Distribution of Variant Prevalence and RR

The objective of this use case was to measure the prevalence of variants. Users may use the strain lineage pie chart by configuring the zip code and time interval of interests. They can then observe variant transition dynamics across space and time by opting for a comparison to visualize the progression pattern categorized by demographics and vaccination status. For example, as shown in Figure 4, the Alpha variant progressed from 3% to 48% and from 0% to 69% in Worcester and Marlborough, respectively, between February and April 2021. A comparative visualization as represented in Figure 4 is not available on MAGGI's user interface. To avoid distraction from an overly complicated pie chart with too many variants, which would be available in MAGGI, we created Figure 4, which only displayed the Alpha variant versus others.

We also calculated the RR of variants for selected geographical regions over time. The RR estimates provide insights on how prevalent a variant is in a specific location compared to the other areas. We used the risk from Central Massachusetts to determine our reference group, and then calculated the RR by dividing the risk from Worcester by the risk from the reference group. For example, we compared the RR of the Delta variant in Worcester with those in other Central Massachusetts regions between May and August 2021 (Table 1). We found that people living in Worcester, the largest city in Central Massachusetts, had a 214% increased risk of Delta infection in May 2021 than those living in other regions in Central Massachusetts. The risk of Delta infection increased rapidly through June and July, and individuals in both Worcester and other Central Massachusetts regions were 20 times more likely to contract the Delta variant in July compared to May 2021. We aimed to probe the social determinants of high RR in the early Delta era at the zip code level, but the small sample size after grouping by zip code and month limited the RR calculation at the zip code level.

Figure 4. Variant tracking from February to April based on Massachusetts' Graphical user interface for Geographic Information (MAGGI) data.

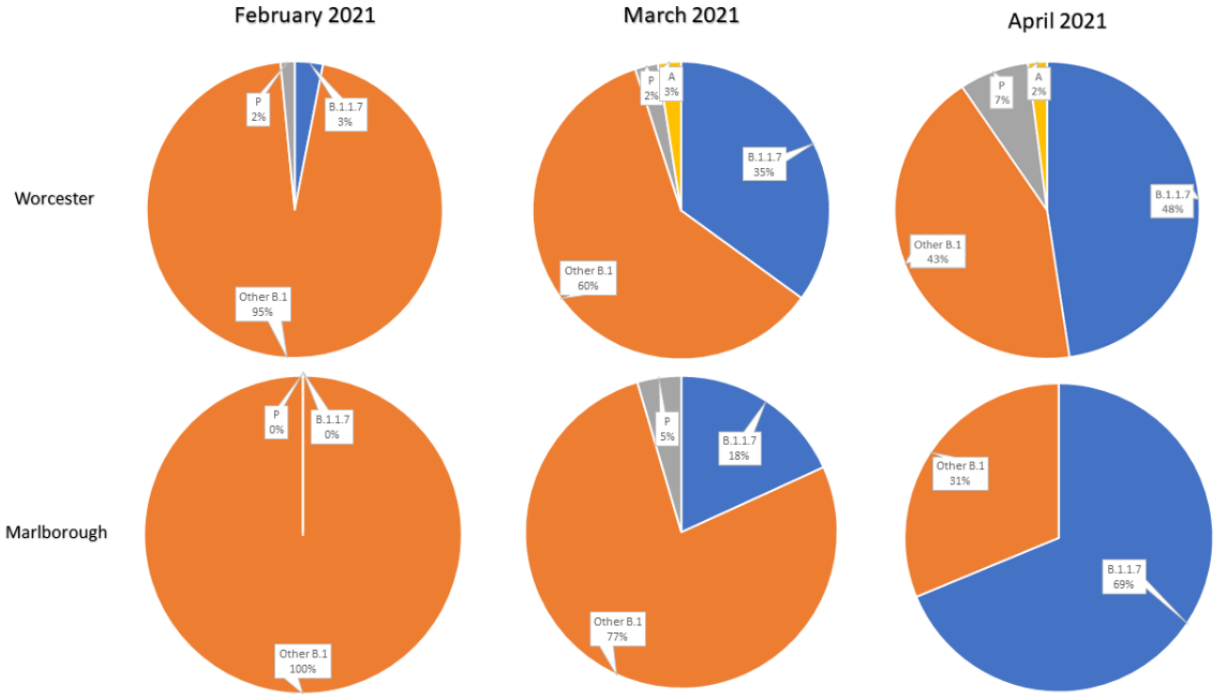


Table 1. Relative risk of infection with the Delta variant from May to August 2021.

Month	Central Massachusetts except Worcester, RR[a] (95% CI)	Worcester, RR (95% CI)
May	Ref[b]	3.14 (2.63-3.64)
June	5.76 (5.13-6.38)	12.75 (11.99-13.51)
July	20.29 (19.86-20.72)	22.56 (21.77-23.35)
August	22.33 (21.90-22.77)	22.34 (21.82-22.87)

[a]RR: relative risk.

[b]Ref: reference group.

Discussion

Principal Findings

We developed MAGGI, a secure platform that enables the spatiotemporal surveillance of the dissemination of SARS-CoV-2 variants. The UMass RDW and ArcGIS Ecosystem provided the foundational tool set for developing this platform. MAGGI is a research tool that is proving to be helpful for research into the socioeconomic and environmental determinants of the COVID-19 pandemic in our local region. Compared to other business intelligence tools, such as Tableau [45], Qlik [46], or Microsoft Power BI [47], the ArcGIS Ecosystem provides more powerful spatial capabilities, such as interactive spatial selection and hot spot cluster analysis. However, the interactive spatiotemporal solution we used can alternatively be developed through a traditional web-mapping approach using JavaScript and open-source spatial databases, such as PostGIS. However, that approach would require several months for a full stack developer and a geospatial team to create and deploy with limited change flexibility due to its resource-intensive coding nature. Although traditional web mapping provides better flexibility in user experience/user interface design, the cost to set up, administer, and support compared to the ArcGIS Ecosystem is greater. Furthermore, we presented our user interface designs in 3 separate meetings to epidemiologists and members of public health agencies, allowing them to provide feedback on the user interface. In the end, they were satisfied with the resulting user interface. For these reasons, we opted to take the ArcGIS route. With the recent development of ArcGIS Online Dashboard, the complexity to develop and deploy an interactive spatial dashboard for disease surveillance or public health management has been significantly reduced. The agility of the whole dashboard creation process enables epidemiologists to promptly determine where and when an infection outbreak occurs and which population it impacts the most.

Our development has multiple strengths. We accrued and sequenced over 5000 clinical samples and engaged genomic epidemiologists to generate questions examining variant transition dynamics in the context of social determinants and vaccination. Currently, the data sets used by the application are updated monthly. Our focus on providing high-resolution GIS is unique and relevant to current times given the COVID-19 pandemic. The Massachusetts DPH is interested in adopting

MAGGI. With the continual emergence of fast-spreading SARS-CoV-2 variants such as Delta and Omicron, GIS technology around COVID-19 will only become more important as this technology advances and the adoption of the technology increases.

Limitations

MAGGI is limited to UMMHC data. Therefore, we only captured patients who got their COVID-19 tests at UMMHC, and MAGGI is more likely to include patients who live near UMass Memorial Hospitals in Central Massachusetts. COVID-19 testing data becomes less accurate as we get further away from the hospital, thus limiting the generalizability of the findings using our tool. We hope to solve this issue by accessing data at the state level. In that case, sequencing data from all hospitals in Massachusetts will populate MAGGI, allowing researchers to make more informed investigations on variants, leading to breaking down the barrier. Another limitation of our tool is the small cohort size that it tracks, with 5000 specimens sequenced at this time.

Future Direction

In the future, we plan to enhance the application in several ways. First, we will be automating data transfer and geoprocessing from the UMass Center for Microbiome Research into the MAGGI system to provide near real-time data analysis with a daily turnaround time instead of monthly. Second, we plan to aggregate street-level data to census geographical units (block groups, tracts) or zip code for the entire MAGGI system and create a public version of this tool so that it can be accessed by the general public. Third, we plan to create ArcGIS StoryMaps, a tool that enables researchers to create interactive narratives around their ideas with a strong sense of place [48]. We identified the ease and efficiency of making an informative presentation using StoryMaps when presenting to Massachusetts DPH officials. With collective feedback from epidemiologists and public officers, we will further polish and enhance the user experience of our tool. Finally, we will add new layers that could expose spatial clusters and hot spots of variants to enrich the geospatial mining ability to detect abnormal regions. Our goal is to enable deep phenotyping analytics using MAGGI to ultimately help our researchers identify actionable interventions.

Acknowledgments

The research reported in this publication was supported by the National Center for Advancing Translational Sciences of the National Institutes of Health under award number UL1-TR001453. However, the content is solely the authors' responsibility and does not necessarily represent the official views of the National Institutes of Health. This study was also supported by the COVID-19 pandemic research fund and institutional funds from the University of Massachusetts Chan Medical School.

Conflicts of Interest

None declared.

References

1. Coronavirus disease (COVID-19) weekly epidemiological update and weekly operational update. World Health Organization. 2020. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports [accessed 2022-06-06]
2. Bernasconi A, Grandi S. A conceptual model for geo-online exploratory data visualization: the case of the COVID-19 pandemic. Information 2021 Feb 06;12(2):69. [doi: 10.3390/info12020069]
3. COVID data tracker. Centers for Disease Control and Prevention. 2022. URL: https://covid.cdc.gov/covid-data-tracker/#datatracker-home [accessed 2022-06-06]
4. Corum J, Zimmer C. Tracking Omicron and other coronavirus variants. New York Times. 2022 May 24. URL: https://www.nytimes.com/interactive/2021/health/coronavirus-variant-tracker.html [accessed 2022-06-06]
5. Coronavirus in the U.S.: latest map and case count. New York Times. 2022 Jun 05. URL: https://www.nytimes.com/interactive/2021/us/covid-cases.html [accessed 2022-06-06]
6. Bauer C, Zhang K, Lee M, Jones M, Rodriguez A, de la Cerda I, et al. Real-time geospatial analysis identifies gaps in COVID-19 vaccination in a minority population. Sci Rep 2021 Sep 13;11(1):18117 [FREE Full text] [doi: 10.1038/s41598-021-97416-y] [Medline: 34518570]
7. Lyseen AK, Nøhr C, Sørensen EM, Gudes O, Geraghty EM, Shaw NT, IMIA Health GIS Working Group. A review and framework for categorizing current research and development in health related Geographical Information Systems (GIS) studies. Yearb Med Inform 2014 Aug 15;9:110-124 [FREE Full text] [doi: 10.15265/IY-2014-0008] [Medline: 25123730]
8. Shneiderman B, Plaisant C, Hesse BW. Improving healthcare with interactive visualization. Computer 2013 May;46(5):58-66. [doi: 10.1109/MC.2013.38]
9. Franch-Pardo I, Desjardins MR, Barea-Navarro I, Cerdà A. A review of GIS methodologies to analyze the dynamics of COVID-19 in the second half of 2020. Trans GIS 2021 Jul 11;25(5):2191-2239 [FREE Full text] [doi: 10.1111/tgis.12792] [Medline: 34512103]
10. Franch-Pardo I, Napoletano BM, Rosete-Verges F, Billa L. Spatial analysis and GIS in the study of COVID-19. a review. Sci Total Environ 2020 Oct 15;739:140033 [FREE Full text] [doi: 10.1016/j.scitotenv.2020.140033] [Medline: 32534320]

11. Iyanda AE, Adeleke R, Lu Y, Osayomi T, Adaralegbe A, Lasode M, et al. A retrospective cross-national examination of COVID-19 outbreak in 175 countries: a multiscale geographically weighted regression analysis (January 11-June 28, 2020). J Infect Public Health 2020 Oct;13(10):1438-1445 [FREE Full text] [doi: 10.1016/j.jiph.2020.07.006] [Medline: 32773211]
12. Sun F, Matthews SA, Yang TC, Hu MH. A spatial analysis of the COVID-19 period prevalence in U.S. counties through June 28, 2020: where geography matters? Ann Epidemiol 2020 Dec;52:54-59.e1 [FREE Full text] [doi: 10.1016/j.annepidem.2020.07.014] [Medline: 32736059]
13. Sannigrahi S, Pilla F, Basu B, Basu AS, Molter A. Examining the association between socio-demographic composition and COVID-19 fatalities in the European region using spatial regression approach. Sustain Cities Soc 2020 Nov;62:102418 [FREE Full text] [doi: 10.1016/j.scs.2020.102418] [Medline: 32834939]
14. COVID-19 GIS Hub. Environmental Systems Research Institute. 2020. URL: https://coronavirus-resources.esri.com/ [accessed 2022-06-06]
15. WHO Coronavirus (COVID-19) Dashboard. World Health Organization. 2022 Jun 06. URL: https://covid19.who.int/ [accessed 2022-06-06]
16. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020 May;20(5):533-534 [FREE Full text] [doi: 10.1016/S1473-3099(20)30120-1] [Medline: 32087114]
17. Zheng S, Edwards JR, Dudeck MA, Patel PR, Wattenmaker L, Mirza M, et al. Building an interactive geospatial visualization application for national health care-associated infection surveillance: development study. JMIR Public Health Surveill 2021 Jul 30;7(7):e23528 [FREE Full text] [doi: 10.2196/23528] [Medline: 34328436]
18. Cahyadi MN, Handayani HH, Warmadewanthi I, Rokhmana CA, Sulistiawan SS, Waloedjo CS, Endroyono, et al. Spatiotemporal analysis for COVID-19 delta variant using GIS-based air parameter and spatial modeling. Int J Environ Res Public Health 2022 Jan 30;19(3):1614 [FREE Full text] [doi: 10.3390/ijerph19031614] [Medline: 35162634]
19. Tang P, Croxen MA, Hasan MR, Hsiao WW, Hoang LM. Infection control in the new age of genomic epidemiology. Am J Infect Control 2017 Feb 01;45(2):170-179 [FREE Full text] [doi: 10.1016/j.ajic.2016.05.015] [Medline: 28159067]
20. Grubaugh ND, Ladner JT, Lemey P, Pybus OG, Rambaut A, Holmes EC, et al. Tracking virus outbreaks in the twenty-first century. Nat Microbiol 2019 Jan 13;4(1):10-19 [FREE Full text] [doi: 10.1038/s41564-018-0296-2] [Medline: 30546099]
21. Chiara M, Horner D, Gissi C, Pesole G. Comparative genomics reveals early emergence and biased spatiotemporal distribution of SARS-CoV-2. Mol Biol Evol 2021 May 19;38(6):2547-2565 [FREE Full text] [doi: 10.1093/molbev/msab049] [Medline: 33605421]
22. Bernasconi A, Mari L, Casagrandi R, Ceri S. Data-driven analysis of amino acid change dynamics timely reveals SARS-CoV-2 variant emergence. Sci Rep 2021 Oct 26;11(1):21068 [FREE Full text] [doi: 10.1038/s41598-021-00496-z] [Medline: 34702903]
23. Yang HC, Chen CH, Wang JH, Liao HC, Yang CT, Chen CW, et al. Analysis of genomic distributions of SARS-CoV-2 reveals a dominant strain type with strong allelic associations. Proc Natl Acad Sci U S A 2020 Dec 01;117(48):30679-30686 [FREE Full text] [doi: 10.1073/pnas.2007840117] [Medline: 33184173]
24. Huang Q, Zhang Q, Bible PW, Liang Q, Zheng F, Wang Y, et al. A new way to trace SARS-CoV-2 variants through weighted network analysis of frequency trajectories of mutations. Front Microbiol 2022 Mar 16;13:859241 [FREE Full text] [doi: 10.3389/fmicb.2022.859241] [Medline: 35369526]
25. Lemieux JE, Siddle KJ, Shaw BM, Loreth C, Schaffner SF, Gladden-Young A, et al. Phylogenetic analysis of SARS-CoV-2 in Boston highlights the impact of superspreading events. Science 2021 Feb 05;371(6529):eabe3261 [FREE Full text] [doi: 10.1126/science.abe3261] [Medline: 33303686]
26. Popa A, Genger JW, Nicholson MD, Penz T, Schmid D, Aberle SW, et al. Genomic epidemiology of superspreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2. Sci Transl Med 2020 Dec 09;12(573):eabj3222 [FREE Full text] [doi: 10.1126/scitranslmed.abe2555] [Medline: 33229462]
27. Argimón S, Abudahab K, Goater R, Fedosejev A, Bhai J, Glasner C, et al. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. Microb Genom 2016 Nov;2(11):e000093 [FREE Full text] [doi: 10.1099/mgen.0.000093] [Medline: 28348833]
28. Theys K, Lemey P, Vandamme AM, Baele G. Advances in visualization tools for phylogenomic and phylodynamic studies of viral diseases. Front Public Health 2019 Aug 2;7:208 [FREE Full text] [doi: 10.3389/fpubh.2019.00208] [Medline: 31428595]
29. Neher RA, Bedford T. Real-time analysis and visualization of pathogen sequence data. J Clin Microbiol 2018 Nov 25;56(11):e00480-e00418 [FREE Full text] [doi: 10.1128/JCM.00480-18] [Medline: 30135232]
30. Norman B, Pedoia V, Majumdar S. Use of 2D U-Net convolutional neural networks for automated cartilage and meniscus segmentation of knee MR imaging data to determine relaxometry and morphometry. Radiology 2018 Jul;288(1):177-185 [FREE Full text] [doi: 10.1148/radiol.2018172322] [Medline: 29584598]
31. Walker RJ, Strom Williams J, Egede LE. Influence of race, ethnicity and social determinants of health on diabetes outcomes. Am J Med Sci 2016 Apr;351(4):366-373 [FREE Full text] [doi: 10.1016/j.amjms.2016.01.008] [Medline: 27079342]
32. Menza T, Hixson L, Lipira L, Drach L. Social determinants of health and care outcomes among people with HIV in the United States. Open Forum Infect Dis 2021 Jul;8(7):ofab330 [FREE Full text] [doi: 10.1093/ofid/ofab330] [Medline: 34307729]

33. Du P, Lemkin A, Kluhsman B, Chen J, Roth RE, MacEachren A, et al. The roles of social domains, behavioral risk, health care resources, and chlamydia in spatial clusters of US cervical cancer mortality: not all the clusters are the same. Cancer Causes Control 2010 Oct 8;21(10):1669-1683. [doi: 10.1007/s10552-010-9596-4] [Medline: 20532608]
34. Kauhl B, Heil J, Hoebe CJPA, Schweikart J, Krafft T, Dukers-Muijrers NHTM. The spatial distribution of hepatitis C virus infections and associated determinants--an application of a geographically weighted Poisson regression for evidence-based screening interventions in hotspots. PLoS One 2015 Sep 9;10(9):e0135656 [FREE Full text] [doi: 10.1371/journal.pone.0135656] [Medline: 26352611]
35. Vermeiren APA, Dukers-Muijrers NHTM, van Loo IHM, Stals F, van Dam DW, Ambergen T, et al. Identification of hidden key hepatitis C populations: an evaluation of screening practices using mixed epidemiological methods. PLoS One 2012 Dec 7;7(12):e51194 [FREE Full text] [doi: 10.1371/journal.pone.0051194] [Medline: 23236452]
36. 2014-2018 American Community Survey 5-year estimates. United States Census Bureau. URL: https://www.census.gov/programs-surveys/acs/technical-documentation/table-and-geography-changes/2018/5-year.html [accessed 2022-06-06]
37. Massachusetts COVID-19 vaccination data and updates. Massachusetts Department of Public Health. URL: https://www.mass.gov/info-details/massachusetts-covid-19-vaccination-data-and-updates [accessed 2022-06-06]
38. CDC/ATSDR Social Vulnerability Index. Agency for Toxic Substances and Disease Registry. 2022 Mar 15. URL: https://www.atsdr.cdc.gov/placeandhealth/svi/index.html [accessed 2022-06-06]
39. Cecret. Github. URL: https://github.com/UPHL-BioNGS/Cecret [accessed 2022-06-06]
40. Epic. URL: https://www.epic.com/ [accessed 2022-06-06]
41. ArcGIS Online. Environmental Systems Research Institute. URL: https://www.esri.com/en-us/arcgis/products/arcgis-online/overview [accessed 2022-06-06]
42. Getis A, Ord JK. The analysis of spatial association by use of distance statistics. In: Anselin L, Rey S, editors. Perspectives on Spatial Data Analysis. Berlin, Heidelberg: Springer; 2010:127-145.
43. Multi-factor authentication (MFA) - Microsoft Security. Microsoft. 2022. URL: https://www.microsoft.com/en-us/security/business/identity-access-management/mfa-multi-factor-authentication [accessed 2022-06-06]
44. ArcGIS Online: share items. Environmental Systems Research Institute. 2022. URL: https://doc.arcgis.com/en/arcgis-online/share-maps/share-items.htm [accessed 2022-06-06]
45. Business intelligence and analytics software. Tableau. URL: https://www.tableau.com/ [accessed 2022-06-06]
46. Qlik. URL: https://www.qlik.com/us/ [accessed 2022-06-06]
47. Data Visualization: Microsoft Power BI. Microsoft. URL: https://powerbi.microsoft.com/en-us/ [accessed 2022-06-06]
48. Health care story maps. Environmental Systems Research Institute. URL: https://storymaps-classic.arcgis.com/en/gallery/#s=30&md=storymaps-industry:health-services [accessed 2022-06-06]

Abbreviations

ESRI: Environmental Systems Research Institute
GIS: Geographic Information System
MAGGI: Massachusetts' Graphical user interface for Geographic Information
RDW: Research Data Warehouse
RR: Relative Risk
UMass: University of Massachusetts
UMMHC: UMass Memorial Health Center
ZCTA: ZIP Code Tabulation Areas

Edited by A Mavragani; submitted 09.03.22; peer-reviewed by A Bernasconi, C Zhao; comments to author 18.04.22; revised version received 08.05.22; accepted 25.05.22; published 13.06.22.

Please cite as:
Shi Q, Herbert C, Ward DV, Simin K, McCormick BA, Ellison III RT, Zai AH
COVID-19 Variant Surveillance and Social Determinants in Central Massachusetts: Development Study
JMIR Form Res 2022;6(6):e37858
URL: https://formative.jmir.org/2022/6/e37858
doi:10.2196/37858
PMID:35658093

Original Paper

General Practitioners' Experiences of Professional Uncertainties Emerging from the Introduction of Video Consultations in General Practice: Qualitative Study

Maja Nordtug[1,2], PhD; Elisabeth Assing Hvidt[3], PhD; Elle Christine Lüchau[3], MA; Anette Grønning[1], PhD

[1]Department for the Study of Culture, University of Southern Denmark, Odense M, Denmark

[2]Department of Social Work, Child Welfare and Social Policy, Oslo Metropolitan University, Oslo, Norway

[3]Department of Public Health, Research Unit for General Practice, University of Southern Denmark, Odense M, Denmark

Corresponding Author:
Maja Nordtug, PhD
Department of Social Work, Child Welfare and Social Policy
Oslo Metropolitan University
Stensberggata 26
Oslo, 0170
Norway
Phone: 47 67236935
Email: majanord@oslomet.no

Abstract

Background: Uncertainties are omnipresent in health care, but little is known about general practitioners' (GPs) professional uncertainties concerning digital consultations. This is problematic, as many countries have undergone an extensive digital transformation.

Objective: The aim of this study was to explore the professional uncertainties that emerged among Danish GPs with the introduction of video consultations.

Methods: We conducted qualitative interviews with 15 Danish GPs during the beginning of the COVID-19 pandemic in 2020. The interviews were analyzed using an abductive approach.

Results: We identified 3 categories of uncertainty: *integrity*, *setting*, and *interaction*. Respectively, these 3 categories of uncertainty refer to (1) uncertainties related to how technology may impede the provision of health care; (2) uncertainties related to the potentials of video technology; and (3) uncertainties related to how the video consultation technology affects interactions with patients.

Conclusions: The uncertainties experienced by Danish GPs appear to be a typical reaction to the introduction of new technology. Embedding video consultation technology into GPs' working routines will take time, and GPs do not necessarily feel intuitively capable of transferring their abilities, such as being good and socially present for video-mediated consultations. The heterogeneity of professional uncertainties experienced among the GPs suggests that they are the product of individual GP-technology relationships—not of the technology in itself. Consequently, we cannot expect that uncertainties can be remedied by changing or precluding new technology.

(JMIR Form Res 2022;6(6):e36289) doi:10.2196/36289

KEYWORDS

video consultation technology; general practice, COVID-19, doctor-patient communication; uncertainties; general practitioners; video consultation; virtual health; physician; digital health; pandemic

Introduction

Background

"Whether a physician is defining a disease, making a diagnosis, selecting a procedure, observing outcomes, assessing probabilities, assigning preferences, or putting it all together" [1], every physician activity can be linked to uncertainty. Uncertainty in health care is a challenge many have acknowledged and made visible, and many conceptual meanings have been attributed to it. In their conceptual taxonomy of the varieties of uncertainty in health care, Han et al [2] defined

uncertainty as fundamentally being "the subjective perception of ignorance" that "makes possible the many manifestations of uncertainty catalogued in conventional definitions of the term—for example, feelings of doubt, perceptions of indefiniteness, indeterminacy, unreliability, and so on." Furthermore, uncertainties in health care vary. According to Han et al [2], the uncertainties can be disease-centered and relate to "diagnosis, prognosis, causal explanations, and treatment recommendation"; they can relate to "structures and processes of care"; or they can be personal and relate to "psychosocial and existential issues."

One aspect of increasing pertinence for health care is assessing the impact of the digital transformation of health care services, such as the implementation of digital consultations, including both email and, most recently, video consultations. Although video consultations are not "straightforward replacements" of physical consultations [3], they are increasingly being used for certain health problems [4]. The important insights about uncertainties in health care mentioned above have not yet been translated to health care professionals' uncertainties related to digital consultations. Consequently, little is known about the role that technologies play in health care professionals' experiences of uncertainty, and even less is known about how feelings of uncertainty concerning digital consultations can be reduced to optimize care and health care professionals' well-being. This is particularly problematic considering how modern health care systems in many countries have undergone a digital transformation in the past decades, which was accelerated during the COVID-19 pandemic. Research shows that technology can either alleviate clinicians' stress or contribute to it, depending on its perceived usability [5,6].

During the COVID-19 pandemic in Denmark, video consultation technology was rapidly implemented in general practice and made available to all patients via the mobile app *Min Læge* ("My Doctor" in Danish) as an alternative to face-to-face consultations. Moreover, Danish health authorities encouraged patients to use video consultations to decrease the transmission of COVID-19 [7,8]. From the beginning of the pandemic in March 2020 to December 2020, 2.6% of all general practice consultations in Denmark were video consultations (not including email and telephone consultations) [9]. The number of video consultations peaked during the first lockdown period in March. However, as society gradually reopened, a rapid decline in video consultations occurred [10]. Although a collective agreement between the Organisation of General Practitioners and Danish Regions suggests and financially supports the continuous use of video consultations in general practice [11], 2.6% suggests a limited adaptation to this new consultation form that might be related to perceived uncertainties among general practitioners (GPs). Research shows that successful adoption of technology depends on continuing use [12]. However, as put forward by Carey and Martin [12], "[un]certainty is a core characteristic associated with new media," and in organizations, employees may not accept the implementation of a new media service even if the organization in question needs it. The introduction and implementation of technology for new patient-provider communication practices, such as video consultations, may also be associated with uncertainties related to new expectations, roles, and working procedures [13-16]. In relation to the recent implementation of video consultations, GPs were found to have mixed experiences, and there is some ambivalence regarding the benefits. Even though some GPs find video consultations to be advantageous, for instance, in that they invite concise and focused consultations, others have indicated disadvantages, such as losing a sense of connection and intimacy with their patients [17]. Furthermore, due to concerns about missing patients' symptoms, some GPs feel that video consultations can compromise the quality of medical work [18].

Objective

To date, few studies have been conducted on GPs' perceived uncertainties related to video consultations, which were rapidly implemented due to the COVID-19 pandemic [19]. With this study, we sought to explore this gap in the research. More specifically, we asked the following question: which professional uncertainties among Danish GPs emerged with the introduction of video consultations in connection with COVID-19?

Exploring this question will provide insights into how the sustained use of video consultations in general practice in Denmark and elsewhere can be optimized, as identifying and categorizing uncertainties can help to address them.

Methods

Data Collection

This study is part of a larger qualitative research project exploring the implications of the use of video consultations as experienced by Danish GPs and patients. The data corpus in the project consists of semistructured interviews with users of video consultations (patients and GPs) and recordings of video consultations. In this paper, we draw on data from interviews with 15 GPs from different clinics: 8 men and 7 women aged 39-61 years from 4 (out of 5) different Danish regions (see Table 1). Of the 15 GPs, 9 had not used video consultations before the onset of the pandemic. Of the GPs who had conducted consultations through video before the pandemic, 5 had used it as persons testing the video technology system, of which 4 had only used it to a very limited extent. The GPs we interviewed were not selected based on any specific inclusion or exclusion criteria, but rather through a convenience sampling technique [20]. This is due to not many GPs having experience with video consultations at the time of our data collection. We were, however, able to secure a variation in the GPs' age and gender and in which region they worked.

Semistructured interviews with the GPs were conducted by ECL and a junior researcher either by phone, video (Zoom, Teams), or at a physical meeting in the clinic. The interviews followed a semistructured interview guide and took place in 2020 during the beginning of the COVID-19 pandemic. The purpose of using a semistructured interview guide was to understand video consultations from the perspective of the GPs [21], allowing room for individual opinions and experiences. Adjustments were made to the interview guide during the period of data collection, as interviewees inspired new interview questions.

Examples of interview questions include the following: "How have you experienced participating in video consultations with patients?"; "What does it mean for your professional relationship with the patient that your consultation takes place via a screen?"; and "Did you feel comfortable letting your consultation take place as a video consultation? If yes/no, what was the reason?" Interviews were audio-recorded and transcribed verbatim by ECL and a student assistant concurrent with the data collection.

Table 1. General practitioner characteristics.

Variable	Our participant sample
Age	39-61 years
Gender	8 men and 7 women
Geographical dispersion	4 out of 5 Danish regions

Data Analysis

The analysis was conducted and developed by the authors of this paper: 3 with a background in media studies and 1 with a background in humanistic health research. All authors have an academic interest in health and media, and 3 have general practice as their clinical research setting and collaborate with GPs who are also researchers. All authors carefully read the transcribed interviews. For the purpose of this paper, we analyzed the data through 2 major stages of coding: first stage and second stage. In the first cycle of coding, researchers initially assign codes to data chunks; in the second cycle of coding, researchers then work with these codes by grouping the codes, such as into categories or themes [22]. MN initially read the interviews and identified various uncertainties through inductive coding using the computer-assisted qualitative data analysis software NVivo (version 12; QSR International). This process served as the first cycle coding of the data [22]. The identified uncertainties were discussed among all authors, and in the second cycle of coding [22], we grouped the data from the coding into 3 general categories. Previous research on uncertainties [2,23] informed our discussions of the data material and our abductive analysis [24]. As such, we went back and forth between the analytical work and previous research, allowing for new insights to emerge from the data while also considering existing research [25]. Through this abductive process, we were able to contribute to the existing research on uncertainties while also not being constrained by previous understandings and classifications. Therefore, we were able to contribute new insights that were built upon previous knowledge.

In our analysis, we focused on GPs' uncertainties in relation to video consultations by focusing on the experiential aspect, as experience "is the place where the mutual relation between human beings and their world can be localized" [26].

Ethics Approval

All patients gave written consent and were informed that participation in the study was voluntary. The study was approved by the institutional board of the University of Southern Denmark and the Research and Innovation Organisation (approval number 10.971) and was conducted in accordance with the General Data Protection Regulation.

Results

Analytical Categories

Below, we present the analytical categories *integrity*, *setting*, and *interaction* using examples that show the nuances and heterogeneities in the uncertainties that the GPs experienced in connection with the introduction of video consultations. The quotes used to exemplify were translated from Danish to English. Each category includes 2 or 3 subcategories, as shown in Table 2.

The categories are not sharply delimited. Rather, they are fluid categories that can help with understanding the variety of the GPs' uncertainties; hence, they are not mutually exclusive. This also means that the categories partially overlap.

Table 2. Categories of professional uncertainties that emerged among Danish general practitioners with the introduction of video consultations.

Category	Subcategory
Integrity	• medical diagnostic • ethical responsibilities • juridical responsibilities
Setting	• off-screen setting • on-screen setting
Interaction	• challenges with focusing and concentrating • challenges with staying present in the web-based environment • challenges with showing emotions in the mediated environment without being able to physically touch each other

Integrity

The category *integrity* comprises the uncertainties that the GPs experienced when having to execute their medical responsibilities through a video consultation. We understand *integrity* in this study as something that "creates the ethical obligation to adhere to standards of intellectual and moral excellence in individual patient care" [27]. The uncertainties relating to integrity revolve around how the use of video consultation technology might impede the provision of health care, potentially bringing the GPs' professionalism into disrepute. Integrity-related uncertainties include the following subcategories: (1) medical diagnostic, (2) ethical responsibilities, and (3) juridical responsibilities.

First, the GPs experienced uncertainties related to medical diagnostics. The GPs expressed how they wished to provide their patients with high-quality care but felt their level of service was compromised due to the use of video consultations. Consequently, some of the GPs expressed that video consultations made them feel "so far away from the GP I want to be" (GP #2). Furthermore, one GP explained, "Just because video has been introduced, we shouldn't create a discounted version of general practice" (GP #3).

In general, the uncertainties relating to medical diagnostics revolved around the GPs feeling sensorily limited due to the video consultation technology only offering the GP a partial view of the patient and precluding the GPs' other senses, including the ability to smell, touch, and feel. Thus, the video consultation technology limited the information available about the patient. The GPs referenced how they would normally get information about the patient "for free" in physical consultations where all their senses could be used. Specifically, the inability to use all senses was associated with situations in which information was retrieved that exceeded the actual medical problem the patients had made an appointment for. One GP explained the following:

> *I can't see how they get up from the waiting room and how they enter the consultation room. I don't see how they huff and puff, because when I talk to them in the video consultation, they are sitting still in their chair.* [GP #5]

Similarly, another GP explained that "sometimes, you also use your sense of smell to see if people smell of alcohol or are well-groomed, or what do I know" (GP #8).

As such, through video consultations, GPs explained that they do not have the same fine-tuned way of sensing their patients as they do in physical consultations. Moreover, some of the GPs expressed uncertainties in connection with the assessment of birthmarks, details of skin eruption, colors, weight, and sore throats through video consultations. These uncertainties appeared to be ascribable to a fear of providing a poorer medical service and inefficient care when consultations were conducted over video. As one GP said, "if you increasingly introduce video consultations in general practice, then you compromise your medical professionalism and you compromise the efficiency" (GP #5).

Second, uncertainties related to integrity also included ethical responsibilities. The GPs came to question their own ethical behavior as a result of having their senses limited. For instance, one GP questioned the ethical appropriateness of asking a patient to film body parts (eg, their thighs and behind) to gather more information about the patient, such as if a patient was suspected to have put on weight: "Well, we might get better at [asking], 'Could you film all of you?' Or…But it's also a little boundary-crossing, you know?" (GP #2).

Similarly, another GP questioned whether patients perceived it as intimidating to get undressed in front of a video camera. The GP described their considerations as follows:

> *It might be that it is easier for them to show—how do you put it?—eczema on their breast, and that it's me who finds it more intimidating that they have to sit in their [own] bathroom and lift up their breast, you know?...They [might] think it's nicer there, than sitting in some man's office taking off their bra.* [GP #9]

Third, uncertainties related to integrity also included whether GPs can legally stand by the assessments they make over video. Not many GPs explicitly expressed these uncertainties, but one GP stated the following:

> *You're behind on points. If something goes wrong...then, naturally, there's a big difference in whether you have seen the patient or not seen the patient from a medico-legal point of view.* [GP #1]

The GP further expressed legal uncertainties about whether seeing a patient through video could be acknowledged as "a proper basis for diagnosing and starting treatments" (GP #1). Consequently, this GP expressed uncertainties about whether the use of video consultations could create legal issues, potentially challenging or risking their professionalism.

Above, we described the 3 subcategories of uncertainties related to integrity. We identified integrity uncertainties among most of the GPs we interviewed. Some, however, expressed that they were not affected by uncertainties related to integrity, as can be seen in the following answer to the question on whether the GP felt professionally safe when diagnosing a patient through a screen: "Yes, I felt comfortable with that" (GP #9). Similarly, another GP answered an interview question on whether they felt sufficiently professional when seeing a patient over video by saying "Yes, I haven't really had any concerns about that" (GP #7).

Some GPs were thus comfortable with making decisions about patients' health after having seen them through a video consultation. Although recognizing that some aspects of patients' health issues might be missed in a video consultation (one GP, for instance, expressed when using video consultations, they missed the last 5% of the total amount of information available about the patient), it was highlighted that "[GPs] could get better at making decisions without [the last 5%]" (GP #12). Thus, the GP suggested that it could be possible to make decisions about a patient's health even if GPs are missing the small percentage of information that the video consultation does not transfer regarding the patient.

Setting

The category *setting* comprises the uncertainties that the GPs experienced in relation to the video technology's potentials. This category relates to the execution of video consultations and includes the following subcategories: (1) off-screen setting and (2) on-screen setting.

First, uncertainties related to off-screen setting emerged because, for many GPs, the video consultation had not become a routine consultation form. Rather, most of the GPs were thrown into the use of video consultations because of the pandemic. Consequently, routines and habits related to video consultations did not exist for patients, GPs, or other health professionals in the GP clinic. One GP questioned how to build a video consultation culture among patients and explained how the patient "also has to sit down, also has to sit in a room without interferences. They can't be on their way to the bakery, they can't be driving their car or..." (GP #5). The GP added that "my own culture around having a video consultation is not very developed" (GP #5).

Another GP explained how hybrid practices (ie, the alternation between onsite and video consultations) were not part of everyday life in the clinic and at times created awkward situations:

> *...it's a bit awkward when I'm sitting and looking in [to the computer], and then a resident comes knocking on my door and "knock knock knock" and "Could you help me with a blood sample?"* [GP #9]

Thus, the GPs were uncertain about how to deal with the off-screen aspects of video consultations.

Uncertainties related to the category *setting* also included on-screen setting uncertainties. Some GPs experienced difficulties with using the computer software when consulting. For example, they considered the new procedure for booking video consultations difficult and found it tricky to look at files on the computer while also having the image of the patient on the screen. Furthermore, some GPs feared that they would unintentionally share their screen: "I might accidentally share some things that I shouldn't share with them [during the video consultation]" (GP #5).

Related to the execution of the consultation, we also identified uncertainties among the interviewed GPs regarding how to adjust their consultation practice to the video medium. One GP talked about their uncertainties regarding how to look at a patient's journal on the computer that was also being used for the video consultation.

> *And the patients, for instance, can't see that as of right now I'm looking in their journal. They can't follow in the same way as if they are sitting here [at the GP's office], then they can see that right now I'm looking at blood samples from the last time [they visited]. If they're sitting at home, they can only see my head and I have to say out loud: "You know what, just a moment. I'll go look," and then you have to remove the image and such.* [GP #9]

In sum, the lack of established culture and practices around video consultations created uncertainties about how to create the best setting for a video consultation.

Interaction

The category *interaction* comprises the uncertainties that emerged in the GPs' interactions with patients in video consultations. Examples of these uncertainties are presented in the following 3 subcategories: (1) challenges with focusing and concentrating, (2) challenges with staying present with the patient in the web-based environment, and (3) challenges with showing emotions in the mediated environment without being able to physically touch each other.

First, challenges with being able to focus and concentrate, especially during longer consultations (eg, talk therapy), were mentioned by several GPs. They stated that they experienced uncertainties related to staying focused and the risk of turning to other tasks during video consultations, such as reading emails. One GP explained the following:

> *And then I also just think it's difficult to concentrate. Basically, you can do all sorts of things at the same time, read emails and reply to things.* [GP #2]

Moreover, the ability to focus and concentrate was experienced as challenging because computers are used for other tasks, such as communication on social media and messages of various kinds (Messenger, Facebook). For example, notifications from other programs that popped up behind the video consultation were perceived as distracting, or as one GP explained: "When you're sitting with your main communication organ in front of you in which all your Facebook thingies appear, then that is constantly disrupting in the background" (GP #4).

Second, regarding the uncertainties connected to the inability to stay present with a patient in the web-based environment, the above-cited GP emphasized that they felt they were not fully present in the video consultation, stating simply, "I'm not as present, I mean" (GP #2).

Furthermore, some, though not all, GPs experienced difficulties with making eye contact with their patients, to which relational importance was accorded. One GP said the following:

> *But I think it's hard for me to have eye contact with the patient. I have to have eye contact with the camera, but I can't see if they...I mean, I have to choose: should they see my eyes, or should I see theirs, you know?* [GP #4]

Using video consultations may cause GPs to feel as if they must choose between looking at their patient and allowing the patient to feel as if they are making eye contact with their GP. However, when asking the GPs whether they experienced the ability to establish and maintain eye contact throughout a video consultation, some GPs answered affirmatively: "Yes, I actually do" (GP #8).

Third, especially when it came to consultations dealing with emotional issues, several GPs experienced uncertainties due to the perceived challenge of showing emotion when being unable to be near or physically touch the patient. For example, one GP stated that "It's difficult to have someone crying or in shock on

the screen. I think the physical—there's a kind of contact [in that situation] that on screen becomes sort of cold and cynical" (GP #15).

Several GPs mentioned that they perceived a lack of potential for action to compensate for the fact that they could not touch the patient. Moreover, one GP also mentioned uncertainties regarding ascertaining a patient's emotional state during a video consultation, questioning whether a patient was about to start crying or just staying silent for a short while. However, the same GP summarized that they found that they had managed this part of video consultations well, mentioning that perhaps this was due to the relational continuity that they were holding with the patient.

Discussion

Principal Findings

Through our analysis, we have contributed new insights on uncertainties in health care by exploring technologically mediated uncertainties as perceived by GPs in Denmark in the period right after the rapid implementation of video consultations in general practice [2,23]. Little has been known about health care professionals' experiences of uncertainty in relation to technology; our study shows some of the challenges GPs experience when new technology is introduced into their working life.

We have identified 3 categories through which we can understand how video consultations create professional uncertainties among some Danish GPs: *integrity*, *setting*, and *interaction*. These categories are closely connected to and mutually shape one another.

In our analysis, the video consultation technology came to play a substantial role in shaping the GPs' experience of their professional uncertainties as having been compromised. The uncertainties related to *integrity* show that the GPs were concerned about whether the video consultation technology would impede the provision of care and, consequently, whether the technology could challenge the GPs' ability to provide quality in health care and thus be the GPs they wanted to be. However, uncertainties are omnipresent in health care, and, as mentioned in the introduction, the implementation of new media is characterized by uncertainty [12]. Historically, the introduction of new communication technologies has led to concerns among GPs. For instance, when the telephone was introduced in general practice, GPs were concerned about the potential extra workload it might create [17,28]. Furthermore, when email consultations were introduced, GPs were found to have concerns about workload, safety, inappropriate use, and confidentiality, as well as difficulties related to diagnosis [17,29]. Consequently, GPs experiencing uncertainties upon the introduction of video consultation technology appears to be a typical reaction.

Technologies are not automatically accepted; they must go through a process until they become normalized, pointing further to the uncertainties related to *setting*. GPs were found to have uncertainties relating to both off-screen and on-screen settings, and the analysis suggests that the practice of video consultation was not yet routinely embedded in general practice [30]—the use of video consultations has not become normalized and well-integrated into the context of general practice consultations. The processes through which a practice becomes routine are complex [30], and when introducing new technology in a profession such as general practice—a profession characterized by rigor and structure [31]—it is to be expected that the implementation process will be complex as well. Thus, adjustment time for GPs to embed the technology into their everyday working life is needed.

For video consultation technology to become routinely embedded in general practice, in addition to not feeling that the technology impedes the provision of health care, GPs must also feel that the interaction through the medium is satisfactory. This points to the uncertainties related to *interaction*. These uncertainties show that some GPs experienced challenges with focusing and concentrating, staying present, and showing emotions. This can be related to the ongoing scholarly discussion about uncertainties that emerge in the digital sphere in general and concerns about the implementation of new technology (telephone, email, and video) in particular. Uncertainty in the digital sphere has been studied for almost 5 decades, focusing on the question of how one perceives other people during a computer-mediated interaction, also termed "social presence" [32,33] and "closeness" [34,35] and related to degrees of "intimacy" [36,37]. The question of to what extent it is possible to perceive presence and closeness through a screen often leads to subjective answers, ranging from distancing to super-connectedness. Hence, an essential point is that users can perceive the same technology's potential for presence and closeness in vastly different ways.

Our analysis also shows that the perceptions and experiences of uncertainties vary, suggesting that uncertainties are not entirely a product of the technology but rather of the GP-technology relationship—a point emphasized at large in science and technology studies [26]. Consequently, different GP technologies are the sources of different uncertainties as reflected in the analysis; for instance, some GPs were comfortable diagnosing patients using video technology, whereas others feared that the quality of their work would be impeded. Similarly, some GPs experienced difficulties with making eye contact with their patients, whereas others did not.

Looking at the above identified GP uncertainties as relationally grounded, the implementation of video consultation in general practice must be acknowledged as "not exclusively a technological issue, but part of a complex organizational change" [38], involving the processes of relating, adapting, or resisting among the whole clinic personnel. In this light, uncertainties cannot necessarily be remedied by changing or precluding the technology but rather by considering the relation between clinic users and the technology when implementing new consultation forms. To ease GPs' uncertainties, future research could delve further into the relations between GPs and technology; the more we know about the relations between GPs and technology, the more we can act on and limit their uncertainties by way of education that may increase feelings of confidence and self-efficacy. However, we also acknowledge that some of the uncertainties experienced by GPs can be a

legitimate reaction to an inappropriate use. Video consultations are not suitable for all health issues, and furthermore, whether a health issue is suitable for video consultation may also be dependent on the individual competences and experiences of the GP and the patient.

Limitations

This study was conducted during an early stage of the introduction of video consultations in general practice; therefore, it cannot state anything about its adaptation over time. Furthermore, as this is a qualitative study, we are unable to make claims about the general frequency of the uncertainties of GPs in Denmark. However, we have no reason to believe that the sample of GPs in this study is not transferrable to other GPs in Denmark.

Acknowledgments

The authors want to thank all study participants who shared their experiences. Furthermore, we wish to acknowledge junior-scholar Nina Primholdt Christensen's contribution in the form of data collection and Helene Lausten Jørgensen for her contribution in the form of transcribing interviews.

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors but was supported by the University of Southern Denmark.

Authors' Contributions

All authors researched literature and conceived the manuscript design. MN wrote the first draft of the manuscript with help from AG. All authors contributed iteratively with adjustments and supplements to the manuscript, discussed the theoretical approach and the findings, and reviewed and approved the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. Eddy DM. Variations in physician practice: the role of uncertainty. Health Aff (Millwood) 1984;3(2):74-89. [doi: 10.1377/hlthaff.3.2.74] [Medline: 6469198]
2. Han PKJ, Klein WMP, Arora NK. Varieties of uncertainty in health care: a conceptual taxonomy. Med Decis Making 2011 Nov;31(6):828-838 [FREE Full text] [doi: 10.1177/0272989x11393976] [Medline: 22067431]
3. Seuren LM, Wherton J, Greenhalgh T, Cameron D, A'Court C, Shaw SE. Physical examinations via video for patients with heart failure: qualitative study using conversation analysis. J Med Internet Res 2020 Feb 20;22(2):e16694 [FREE Full text] [doi: 10.2196/16694] [Medline: 32130133]
4. Howgego G, Sharma I, Kalu P. The rules for online clinical engagement in the COVID era. J Plast Reconstr Aesthet Surg 2020 Dec;73(12):2127-2135 [FREE Full text] [doi: 10.1016/j.bjps.2020.08.045] [Medline: 33051176]
5. Kaihlanen AM, Gluschkoff K, Hyppönen H, Kaipio J, Puttonen S, Vehko T, et al. The associations of electronic health record usability and user age with stress and cognitive failures among Finnish registered nurses: cross-sectional study. JMIR Med Inform 2020 Nov 18;8(11):e23623. [doi: 10.2196/23623] [Medline: 33206050]
6. Poon EG, Rosenbloom ST, Zheng K. Health information technology and clinician burnout: current understanding, emerging solutions, and future directions. J Am Med Inform Assoc 2021 Apr 23;28(5):895-898 [FREE Full text] [doi: 10.1093/jamia/ocab058] [Medline: 33871016]
7. Videokonsultationer skal mindske smitterisikoen hos de praktiserende læger. Video consultations to minimize the risk of infection in general practice. Praktiserende Lægers Organisation. 2020 Mar 17. URL: https://www.laeger.dk/nyhed/videokonsultationer-skal-mindske-smitterisikoen-hos-de-praktiserende-laeger [accessed 2021-04-11]
8. Lüchau EC, Jepsen C, Grønning A, Assing Hvidt E. Reciprocal dynamics between patients' choice of place and how they experience video consultations: a qualitative study. Digit Health 2021 Oct 28;7:20552076211052158 [FREE Full text] [doi: 10.1177/20552076211052158] [Medline: 34733540]
9. Et digitalt år med corona. A digital year with Corona. Danske Regioner. 2021 Mar 22. URL: https://www.regioner.dk/services/nyheder/2021/marts/et-aar-med-corona [accessed 2021-04-11]
10. COVID-19: Monitorering af aktivitet i sundhedsvæsenet: Beskrivelse af udviklingen i aktivitet i sundhedsvæsenet under COVID-19 epidemien – 10. rapport. COVID-19: monitoring activity in the health care system: description of the development in activity in the health care system during the COVID-19 epidemic - 10th report. Sundhedsstyrelsen. 2021 Jun. URL: https://www.sst.dk/-/media/Udgivelser/2020/Corona/Monitorering/10_-monitoreringsrapport.ashx [accessed 2022-06-07]
11. Regionernes lønnings- og takstnævn. The wages and charge board of the regions. Praktiserende Lægers Organisation. 2021 Jul 05. URL: https://www.laeger.dk/sites/default/files/forhandlingsaftale_ok22.pdf [accessed 2021-12-13]

12. Carey J, Martin CJE. When media are new: understanding the dynamics of new media adoption and use. Ann Arbor, MI: University of Michigan Press; 2010.
13. Assing Hvidt E, Grønning A, Nisbeth Brøgger M, Møller JE, Fage-Butler A. Multilevel structures and human agency in relation to email consultations: a strong structuration theory analysis of the Danish general practice setting. Soc Sci Med 2021 Aug;282:114155 [FREE Full text] [doi: 10.1016/j.socscimed.2021.114155] [Medline: 34174578]
14. Greenhalgh T, Stones R, Swinglehurst D. Choose and Book: a sociological analysis of 'resistance' to an expert system. Soc Sci Med 2014 Mar;104:210-219 [FREE Full text] [doi: 10.1016/j.socscimed.2013.12.014] [Medline: 24581080]
15. Grønning A, Assing Hvidt E, Nisbeth Brøgger M, Fage-Butler A. How do patients and general practitioners in Denmark perceive the communicative advantages and disadvantages of access via email consultations? a media-theoretical qualitative study. BMJ Open 2020 Oct 26;10(10):e039442 [FREE Full text] [doi: 10.1136/bmjopen-2020-039442] [Medline: 33109668]
16. Assing Hvidt E, Søndergaard J, Klausen M, Grønning A. Not just an information-delivery tool. an ethnographic study exploring Danish GPs' perspectives on and experiences with the relational potential of email consultation. Scand J Prim Health Care 2020 Dec;38(4):411-420 [FREE Full text] [doi: 10.1080/02813432.2020.1843939] [Medline: 33185136]
17. Due TD, Thorsen T, Andersen JH. Use of alternative consultation forms in Danish general practice in the initial phase of the COVID-19 pandemic - a qualitative study. BMC Fam Pract 2021 Jun 02;22(1):108. [doi: 10.1186/s12875-021-01468-y] [Medline: 34078281]
18. Grønning A, Assing Hvidt E. Den digitale læge-patient-relation. The digital doctor-patient relation. In: Assing Hvidt E, Søndergaard J, Grønning A, editors. Relationer i Sundhedsvæsenet: hvorfor og hvordan? Relations in the health care system: why and how?. Frederiksberg, Denmark: Samfundslitteratur; Aug 17, 2021:249-269.
19. Randhawa RS, Chandan JS, Thomas T, Singh S. An exploration of the attitudes and views of general practitioners on the use of video consultations in a primary healthcare setting: a qualitative pilot study. Prim Health Care Res Dev 2019 Jan;20:e5 [FREE Full text] [doi: 10.1017/S1463423618000361] [Medline: 29909798]
20. Paik LS, Shahani-Denning C. Convenience Sampling. In: Rogelberg SG, editor. The SAGE Encyclopedia of Industrial and Organizational Psychology. 2nd ed. Thousand Oaks, CA: Sage; 2017.
21. Kvale S, Brinkmann S. InterViews: Learning the craft of qualitative research interviewing. 2nd ed. Thousand Oaks, CA: Sage; 2009:0761925422.
22. Miles MB, Huberman AM, Saldaña J. Qualitative Data Analysis. 3rd ed. Thousand Oaks, CA: Sage; 2014.
23. Pomare C, Churruca K, Ellis LA, Long JC, Braithwaite J. A revised model of uncertainty in complex healthcare settings: a scoping review. J Eval Clin Pract 2019 Apr;25(2):176-182. [doi: 10.1111/jep.13079] [Medline: 30467915]
24. Timmermans S, Tavory I. Theory construction in qualitative research: from grounded theory to abductive analysis. Sociological Theory 2012 Sep 10;30(3):167-186. [doi: 10.1177/0735275112457914]
25. Inversen SM. Med NVivo som analyseparter. Using NVivo as a partner in analysis. In: Drotner K, Iversen SM, editors. Digitale metoder. At skabe, analysere og dele data. Digital methods. To create, analyze, and share data. Frederiksberg, Denmark: Samfundslitteratur; Aug 2017:147-163.
26. Verbeek PPCC. Don Ihde: the technological lifeworld. In: Achterhuis HJ, editor. American Philosophy of Technology: The Empirical Turn. Bloomington, IN: Indiana University Press; 2001:119-146.
27. Chervenak FA, Grünebaum A, Bornstein E, Wasden S, Katz A, Rochelson BL, et al. Expanding the concept of the professional integrity of obstetrics during a public health emergency. J Perinat Med 2020 Jun 25;48(5):435-437. [doi: 10.1515/jpm-2020-0174] [Medline: 32374289]
28. Lee SA, Zuercher RJ. A current review of doctor–patient computer-mediated communication. J Commun Healthc 2017 Jan 31;10(1):22-30. [doi: 10.1080/17538068.2017.1282184]
29. DeVille K, Fitzpatrick J. Ready or not, here it comes: the legal, ethical, and clinical implications of E-mail communications. Semin Pediatr Surg 2000 Feb;9(1):24-34. [doi: 10.1016/s1055-8586(00)70005-9] [Medline: 10688383]
30. May C, Finch T. Implementing, embedding, and integrating practices: an outline of normalization process theory. Sociology 2009 Jun 15;43(3):535-554. [doi: 10.1177/0038038509103208]
31. Kurtz S, Silverman JD. The Calgary-Cambridge Referenced Observation Guides: an aid to defining the curriculum and organizing the teaching in communication training programmes. Med Educ 1996 Mar;30(2):83-89 [FREE Full text] [doi: 10.1111/j.1365-2923.1996.tb00724.x] [Medline: 8736242]
32. Rice RE. Contexts of research on organizational computer-mediated communication. a recursive review. In: Lea M, editor. Contexts of Computer-Mediated Communication. Hertfordshire, UK: Harvester Wheatsheaf; 1992:113-144.
33. Short J, Williams E, Christie B. The Social Psychology of Telecommunications. Hoboken, NJ: Wiley; 1976.
34. Kim D, Frank MG, Kim ST. Emotional display behavior in different forms of computer mediated communication. Comput Human Behav 2014 Jan;30:222-229 [FREE Full text] [doi: 10.1016/j.chb.2013.09.001]
35. Riordan MA, Kreuz RJ. Cues in computer-mediated communication: a corpus analysis. Comput Human Behav 2010 Nov;26(6):1806-1817 [FREE Full text] [doi: 10.1016/j.chb.2010.07.008]
36. Feenberg A. The written world: on the theory and practice of computer conferencing. In: Mason R, Kaye A, editors. Mindweave: communication, computers and distance education. Oxford, UK: Pergamon Press; 1989:22-39.
37. Moran C, Hawisher GE. The rhetorics and languages of electronic mail. In: Snyder I, editor. Page to Screen: Taking Literacy into the Electronic Era. London, UK: Routledge; 1998:80-101.

38. Giunti G, Goossens R, De Bont A, Visser JJ, Mulder M, Schuit SCE. The need for sustainable teleconsultation systems in the aftermath of the first COVID-19 wave. J Med Internet Res 2020 Oct 05;22(10):e21211 [FREE Full text] [doi: 10.2196/21211] [Medline: 32997642]

Abbreviations

GP: general practitioner

Edited by A Mavragani; submitted 09.01.22; peer-reviewed by A Hidki, T Skonnord, S Pesälä; comments to author 06.04.22; revised version received 26.04.22; accepted 04.05.22; published 14.06.22.

Please cite as:
Nordtug M, Assing Hvidt E, Lüchau EC, Grønning A
General Practitioners' Experiences of Professional Uncertainties Emerging from the Introduction of Video Consultations in General Practice: Qualitative Study
JMIR Form Res 2022;6(6):e36289
URL: https://formative.jmir.org/2022/6/e36289
doi:10.2196/36289
PMID:35653607

Original Paper

Integration vs Collaborative Redesign Strategies of Health Systems' Supply Chains in the Post-COVID-19 New Normal: Cross-sectional Survey Across the United States

Jiban Khuntia[1], PhD; Frances J Mejia[1], BSc; Xue Ning[2], PhD; Jeff Helton[1], PhD; Rulon Stacey[1], PhD

[1]Health Administration Research Consortium, University of Colorado Denver, Denver, CO, United States

[2]Business Department, University of Wisconsin Parkside, Kenosha, WI, United States

Corresponding Author:
Jiban Khuntia, PhD
Health Administration Research Consortium
University of Colorado Denver
1475 Lawrence St.
Denver, CO, 80202
United States
Phone: 1 303 315 8424
Email: jiban.khuntia@ucdenver.edu

Abstract

Background: Given the widespread disruptions to supply chains in 2020 because of the COVID-19 pandemic, questions such as how health systems are shaping strategies to restore the supply chain disruptions are essential to have confidence in health systems' supply chain model strategies. Plausibly, health systems have an opportunity for redesign, growth, and innovation by utilizing collaborative strategies now, compared to the usual strategies of integrating their existing supply chains to reduce inefficiencies.

Objective: This study focuses on teasing out the nuance of supply chain integration versus collaborative redesign strategies for health systems in the post-COVID-19 new normal. We focus on 2 research questions. First, we explore the impact of perceived supply chain challenges and disruptions on health systems' supply chain integration (SC-INTEGRATION) and collaborative redesign (SC-REDESIGN) strategies. Second, we examine the outcomes of integration and collaborative redesign strategic choices on growth and service outcomes.

Methods: We used data for this study collected through a consultant from a robust group of health system chief executive officers (CEOs) across the United States from February to March 2021. Among the 625 health system CEOs contacted, 135 (21.6%) responded to our survey. We considered supply chain–relevant strategy and outcome variables from the literature and ratified them via expert consensus. We collected secondary data from the Agency for Healthcare Research and Quality (AHRQ) Compendium of the US Health Systems, leading to a matched data set from the 124 health systems. Next, we used ordered logit model estimation to examine CEO preferences for partnership strategies to address current supply disruptions and the outcomes of strategy choices.

Results: Health systems with higher disruptions would choose integration (positive, $P<.001$) over redesign, indicating that they still trust the existing partners. Integration strategy is perceived to result in better service outcomes ($P<.01$), while collaborations are perceived to lead to greater growth opportunities ($P<.05$); however, the role of integration in growth is not entirely ruled out (combined model, $P<.001$). Plausibly, some health systems would choose integration and collaborative redesign models, which have a significant relationship with both services (combined model, $P<.01$) and growth, establishing the importance of mixed strategies for health systems.

Conclusions: The cost of health care continues to rise, and supply-related costs constitute a large portion of a hospital's expenditure. Understanding supply chain strategic choices are essential for a health system's success. Although collaboration is an option, focusing on and improving existing integration dynamics is helpful to foster both growth and services for health systems.

(JMIR Form Res 2022;6(6):e35317) doi:10.2196/35317

KEYWORDS

COVID-19; post-COVID-19; health systems; supply chain integration; supply chain collaboration; supply chain resilience

Introduction

Background

Supply chain disruptions during the COVID-19 pandemic presented multiple challenges for health systems. These included delays in several modes of transportation and delays in and inability to acquire critical supplies, leading to low inventories, backup suppliers sourcing from the same source as a critical supplier, and critical suppliers going into liquidation. A survey indicated that 10%-18% of health care organizations reported a significant impact, 4.1%-16% reported a severe impact, and 3%-5% reported a catastrophic impact on their businesses [1]. Supply chain disruptions adversely impacted operations, emergency plans, and responses during the pandemic and need to be revamped to improve health systems' functions [2].

After the pandemic, health systems are in the process of restoring supply chains to be resilient. However, not all are blaming their current partners, as some of the challenges were beyond and above the scope of the supply chain delivery models—such as overly dependency on low-cost global value chains or lack of availability of alternative local vendors. Nevertheless, the disruptions invoked new efforts to at least focus on the supply chain resiliency and, accordingly, explore ways to focus on integration or redesign of supply chains along with other aspects [3]. It is essential to explore what factors influence health systems to choose between integration or redesign strategies and which of these strategies are effective for health system outcomes. The choice of either integration or collaboration strategies may be consequential or pose different challenges.

Prior research notes that uncooperative existing supply chain partners may lead to less cost-effective health outcomes or fail in implementing standardized health care processes [4,5]. In addition, redesigning a supply chain is not easy, as the complexities of supply chain management in the health care sector are high [4,6,7]. Nevertheless, since supply-related costs constitute 30%-50% of a hospital's operating expenses [8], integrating or redesigning the supply chain is crucial for any health system.

The tension between adopting either integration or redesign is exacerbated postpandemic. Integration strategies adopted to cope with the supply chain challenges of the pandemic will require evaluation and modification to be sustainable over a longer-term, stable market landscape. Further, the compatibility of organizations that integrated during the pandemic should be evaluated to understand how sourcing strategies, pricing, and shipping logistics work on both supplier and user sides of the relationship [9,10].

Health systems appear to be involved in—or considering—integration with suppliers or redesigning supply chain relationships based on anecdotes and thorough responses to our survey. Thus, this study will focus on 2 research questions arising from these preliminary observations. First, we examine the impact of supply chain disruptions on health systems' supply chain integration and redesign strategies. Second, we examine whether integration or redesign has a better outcome on growth and services.

Findings suggest that disruptions and challenges influence supply chain partnership choice. Higher disruptions tend to lead to integrative strategies. This finding could imply a greater trust among existing partners or the preference to avoid further burdening a complex system by adding more collaborative arrangements. Additionally, although collaboration provides greater growth opportunities, integration could also facilitate growth. The improvements in service delivery could provide resource allocation opportunities for organizations, allowing a shift in focus to growth initiatives.

Examining the supply chain strategies in health systems when facing challenges and the consequent outcomes will guide new supply chain management strategies and health care policies. Policy implementation to mitigate the financial risk associated with collaborative initiatives may impact partnership choice. Incentives to choose supply chain redesign despite its added complexity facilitate growth opportunities. Supply chain redesign through alternative sourcing facilities, for example, fosters resilience and provides sustainable solutions to supply chain disruptions.

Literature Review: Integration or Redesign of Supply Chains

Operational challenges and supply chain inefficiencies have driven organizations to foster resilience through supply chain integration and redesign. The complexities of supply chain operations found within and between nodes contribute to supply chain disruptions. To integrate current systems to develop lean and agile supply chain operations, some organizations faced greater challenges due to the disruptions brought on by the COVID-19 pandemic. This has highlighted the importance of continual supply chain evaluations and the implementation of proactive strategies to foster resilience. This literature review examines the factors contributing to and the severity of supply chain disruptions. Next, we examine organizational resilience and how organizations have fostered resilience and have found supply chain sustainability solutions through supply chain integration and redesign.

Supply chain inefficiencies and disruptions present unique challenges for health systems. Factors specific to the health sector contribute to these challenges more. For instance, physicians are key decision makers in the procurement of prescription drugs but may have limited understanding of the production and supply chain or may think it is different from other sectors [4]. The health care sector is under regulatory pressures and long drug developmental cycles, and inventory management complexities in predicting a patient mix and supply consumption are added challenges [6]. Further, to maximize the benefit of supply chain operations, management must

improve supplier relationship management, logistics operational tools, and process improvement, similar to other sectors.

The COVID-19 pandemic has exacerbated issues faced in supply chain operations. Some health care supply chains focus on minimizing costs by following a lean, agile inventory approach; however, during the pandemic, there was a pause in the movement of materials, causing massive supply chain disruptions. During the COVID-19 pandemic, health care providers' personal protective equipment (PPE), medical-surgical supplies, and pharmaceuticals faced supply chain disruptions. Hospitals' efforts to mitigate and minimize potential supply chain disruptions during COVID-19 were unsuccessful due to the overreliance on overseas manufacturing [11,12].

Health care organizations operate in disruptive environments. Continual evaluation of supply chain operations and proactive actions must be implemented to minimize critical supply shortages. Prior research points to developing and updating robust continuity of the supply chain through various means, such as an agile and innovative culture, communication practices, business continuity, and sourcing strategies [11-13]. However, in practice, a hospital may choose to engage in a spectrum of supply-related management, ranging from complete internal control to complete outsourcing. In this process, factors such as product or service characteristics, spatial complexity, degree of goal congruence, regulatory environment, and physical characteristics of the health system are vital to determine the strategic choices [6].

The processes and determinants of supply chain disruption severity include density, complexity, and node criticality [14]. Density is the quantity and geographic spacing of nodes in a supply chain; highly dense areas are more likely to experience significant disruptions if there is a significant portion of sourcing from those impacted areas. The complexity here refers to the relationship between the number of nodes in a supply chain and the number of connections among the nodes. On the one hand, a less complex system with fewer nodes and connections may experience less significant disruptions, but on the other hand, the presence of extra nodes can act as a buffer for supplier setbacks and can thus increase resiliency despite the added complexity. Node criticality is the relative importance of a node, and supply chains with a more significant number of critical nodes have greater probabilities of experiencing disruptions than supply chains in which critical processes are distributed or shared among several nodes. Supply chain integration drives optimal operations through the seamless exchange of information and the flow of products. Successful integration can increase organizational competitiveness through increased flexibility and responsiveness [15]. This change in operational strategy is standard and increases an organization's capability to provide services.

The challenges of inefficient supply chains can drive organizations to foster resilience and overcome these challenges through supply chain integration and collaboration. Resilience can be defined as the ability of a supply chain system to reduce the probability of disruptions, reduce the consequences of disruptions, and reduce recovery time from disruptions back to normal operations [14]. Thus, supply chain resiliency starts with the organizations' ability to proactively address unanticipated events. Suppliers can significantly impact the production of critical supplies used by health organizations, as seen by the COVID-19 pandemic [16].

Many organizations utilize information systems and integrate information technology into their supply chain operations to improve integration through infrastructural support, value creation, logistic operations, and supply chain management performance [16-18]. The utilization of information technology to optimize health care delivery includes telehealth, electronic medical records, and the implementation of decision support systems. Additionally, health care organizations have leveraged innovations by utilizing cloud-based information sharing to improve supply chain visibility and improve the system's responsiveness to market forces [19]. Although associated privacy issues remain a significant concern in outsourcing health care data, privacy-preserving hybrid frameworks have been proposed to transit insensitive data to commercial cloud systems and the rest to trusted private cloud systems [20]; newer technologies, such as blockchains, provide several of the trusting and transparency-based tracking functions [21,22].

The use of cloud-based information sharing in hospitals is better positioned to efficiently respond to patient demand, reducing inventory costs, supply costs, and supply shortages. Radio-frequency identification (RFID) also improves inventory visibility at various supply chain stages to reduce shrinkage and shipping errors [5]. Supply chain redesign through start-ups or other entrepreneurial initiatives involves the collaboration of 2 or more organizations to execute supply chain operations, of which partnerships can be ongoing or limited.

Collaboration can be further divided into vertical and horizontal components. Vertical collaboration can include collaborations with customers, internal functions, and suppliers, while horizontal collaboration includes relationships with competitors and noncompetitors [23]. Supply chain collaboration is less chosen, possibly due to its difficulty to implement, overreliance on technology in trying to implement it, failure to differentiate between whom to collaborate with in the segmentation of customers and suppliers, a fundamental lack of trust between trading partners, and inequality between and among partners [23,24]. Despite collaboration challenges, these partnerships can foster organizational growth through innovation. Interconnecting elements in collaboration include information sharing, goal congruence, decision synchronization, incentive alignment, resource sharing, collaborative communication, and joint knowledge creation [25].

A successful collaboration leverages the knowledge and resources of suppliers and buyers. Large organizations may enter partnerships with start-ups because of the creativity, flexibility, and agility; large organizations can offer financial resources and execution of these ideas [26]. In 1 collaborative effort, physician leaders and stakeholder groups launched clinical communities focused on specialized care, including joint replacement, spine surgery, and blood management. The partnership between the supply chain team and lead physicians addressed the growing cost of health care through supply cost

initiatives, realizing significant cost savings. The combined efforts of this initiative, which included medical supply consolidation, resulted in multimillion-dollar savings [27].

The current COVID-19 pandemic has brought to light weaknesses in the health care supply chain, resulting in delays or backorders of supply requisitions. The potential issues with delivery delays and backorders are myriad but include impacts such as duplicate orders or multiple orders to different vendors, resulting in excessive inventory holdings and poor working capital management. Cooperation between suppliers and buyers requires frequent communication on order status, shipping delays, coordination of substitute products as needed, and fair allocation of available stock among multiple buyers. Cooperation between these parties can also identify potential causes of backorders and shipping delays that ultimately improve the efficiency of product movement through the supply chain and optimize investments in inventory [28].

A common way to synthesize these multiple supply chain challenges is through a group purchasing organization (GPO), which represents a collaboration between suppliers, distributors, and end-user organizations to reduce costs and increase the efficiency of supply chain operations [8]. Although a GPO can reduce supply costs to health systems, it requires some compromises in standardizing item usage (eg, exam gloves or intravenous [IV] sets) to increase the bargaining power that yields such a low price. Many of the collaborative models noted earlier in this review have elements of cooperation between suppliers and buyers without the price negotiation dynamics. Nevertheless, successful integration and redesign strategies offer organizations service and growth opportunities. Supply chain integration has led to service improvements within organizations. The need for agility in supply chain operations has led organizations to leverage their integrated capabilities to create a more seamless exchange of information. Other organizations have leveraged collaborations to capitalize on business growth opportunities. Increased collaborative efforts by adding manufacturing nodes have allowed organizations to address shortage issues. Global supply chain sourcing was disrupted, facilitating the inclusion and collaboration of local sourcing facilities [29]. In this context, investigating the impact of supply chain disruptions on health systems' supply chain integration and redesign strategies and whether integration or redesign has a better outcome on growth and services is timely and informative to both practice and research.

Methods

Data Collection

The effort to assess the linkage between the competition and integration prospects of health systems is part of a broad project undertaken by the Health Administration Research Consortium at the Business School of the University of Colorado Denver [30]. The project involves an annual and broad study on health systems and collects insights via a survey of health system chief executive officers (CEOs).

Data for this study were collected in collaboration with a consultant from a robust group of health system CEOs across the United States from January to March 2021. Expert inputs were taken, and the survey was validated and pilot-tested with 5 top executives from the Health Administration Program Advisory Board. The survey questionnaire was revised and finalized in January 2021. The specific questions that were asked in the survey instrument to measure the supply chain variables are as follows:

- I believe that the most pressing issue facing the growth of my health system in 2021 is supply chain disruptions and challenges.
- Are you currently engaged with or considering engaging with any of the following types of partners through joint ventures, strategic alliances, or informal collaborations to support your growth? The second question has the options of (1) supply chain and logistics organizations, (2) start-ups or entrepreneurial collaborations, (3) the new normal is presenting growth opportunities different than before COVID-19, (4) health delivery and services overall will improve over the next 12 months.

These questions used a 7-point Likert scale that varied from 1 = strongly disagree to 7 = strongly agree.

A contact list of CEOs was compiled from 624 health systems across the United States using data from multiple sources, contacts, professional networks, websites, and annual reports. The survey instrument was implemented in a professional survey platform and was mapped with emails to the platform to create unique, trackable links for each health system. Email and phone solicitations were made in multiple rounds between January 25 and March 2, 2021. A total of 148 responses were received from the 624 CEOs contacted, representing a 23.7% response rate, of which 13 (8.8%) incomplete responses could not be used, leaving 135 (21.6%) final usable responses. The 135 health systems represented in this survey varied from 1 to 18 hospitals with 176-75,000 employees. The annual revenue in 2020 of the health systems ranged from US $0.7 million to US $14 billion. The health systems aggregately represented US $300 billion in revenues and 1.1 million employees across the United States.

We then matched the survey data set with the secondary data collected from the Agency for Healthcare Research and Quality (AHRQ) *Compendium of the US Health Systems* to understand a better and complete picture of the health systems. Finally, we obtained data from 124 health systems located across the United States. We analyzed this combined data set to report several insights in this study.

Ethics Consideration

An ethics review was not applicable for this study. The data used was received through a leading professional consulting firm that anonymizes and provides secondary firm-level data for research and analysis to draw insights.

Variables and Measures

Table 1 shows the description of supply chain variables used in this study. The main variables in this study are supply chain disruptions and challenges (SC-DISRUP), presenting growth opportunities in the post-COVID-19 new normal (GROWTH), overall improvements in health delivery and services

(SERVICE-IMPR), integration with supply chain and logistics organizations (SC-INTEGR), and redesign through start-ups or entrepreneurial collaborations (SC-REDEGN).

The influencing factors examined in this study are in several categories: size, region, teaching status, revenue, and several other system characteristics. These variables were coded in the way shown in Table 2 to reflect the characteristics of a health system, which may influence its supply chain disruptions, challenges, and growth in the post-COVID-19 new normal. In summary, 3 size variables measured the number of beds across a health system (SIZE_B-SMALL, SIZE_B-MEDIUM, SIZE_B-LARGE), 4 region variables reflected the location of a health system (REGION-NE, REGION-MW, REGION-SOUTH, REGION-WEST, where NE refers to "Northeast" and MW to "Midwest"), 3 variables were related to teaching status (TEACHING-NON, TEACHING-MINOR, TEACHING-MAJOR), and 3 revenue variables measured the annual revenue of a health system (REVENUE-LOW, REVENUE-MEDIUM, REVENUE-HIGH), in addition to variables about the disproportionate share hospital (DSH) patient (HIGH-DSH-HOSP), uncompensated care burden (HIGH-BURDEN-SYS and HIGH-BURDEN-HOSP), ownership status (OWNERSHIP), number of physicians (PHYSICIANS), and number of hospitals (HOSPITALS).

Table 1. Description of supply chain variables.

Variable	Description
SC-DISRUP[a]	Supply chain disruptions and challenges
SC-INTEGR[b]	Partner integration: supply chain and logistics organizations
SC-REDEGN[c]	Partner redesign: start-ups or entrepreneurial collaborations
GROWTH[d]	New normal: presenting growth opportunities
SERVICE-IMPR[e]	New normal: health delivery and services overall will improve

[a]SC-DISRUP: supply chain disruptions and challenges.

[b]SC-INTEGR: integration with supply chain and logistics organizations.

[c]SC-REDEGN: redesign through start-ups or entrepreneurial collaborations.

[d]GROWTH: presenting growth opportunities in the post-COVID-19 new normal.

[e]SERVICE-IMPR: overall improvements in health delivery and services.

Table 2. Description of contingent variables.

Variable	Description
Size: The 3 size variables of the health system were coded using the total beds managed by the health system across all hospitals, reported by the AHRQ[a]*Compendium of the US Health Systems*.	
SIZE_B-SMALL	The health system has <100 beds.
SIZE_B-MEDIUM	The health system has 100-400 beds.
SIZE_B-LARGE	The health system has >400 beds.
Region: The 4 region variables of the health systems were coded based on their primary location in the United States, following the Census Bureau categorization.	
REGION-NE[b]	The health system in the Northeast
REGION-MW[c]	The health system in Midwest
REGION-SOUTH	The health system in the South
REGION-WEST	The health system in the West
Teaching: The 3 teaching variables were coded based on the teaching status of a health system.	
TEACHING-NON	A nonteaching health system
TEACHING-MINOR	A minor teaching health system
TEACHING-MAJOR	A major teaching health system
Revenue: The 3 revenue variables of the health systems were coded using the annual revenue of the health system across all hospitals.	
REVENUE-LOW	Revenue<US $2 billion
REVENUE-MEDIUM	Revenue=US $2-5 billion
REVENUE-HIGH	Revenue>US $5 billion
HIGH-DSH[d]-HOSP	The health system includes at least 1 high-DSH-patient-percentage hospital: 1=yes, 0=no.
HIGH-BURDEN-SYS	Health system–wide uncompensated care burden flag: 1=yes, 0=no
HIGH-BURDEN-HOSP	The health system includes at least 1 high-uncompensated-care-burden hospital: 1=yes, 0=no.
OWNERSHIP	Predominantly investor-owned hospitals: 1=yes, 0=no
PHYSICIANS	The number of physicians in the health system is measured by the number of physicians reported by the AHRQ *Compendium of the US Health Systems*.
HOSPITALS	This is the number of hospitals the health system has, as reported by the AHRQ *Compendium of the US Health Systems*.

[a]AHRQ: Agency for Healthcare Research and Quality.

[b]NE: Northeast.

[c]MW: Midwest.

[d]DSH: disproportionate share hospital.

Statistical Analysis

To answer the 2 research questions, we had 2 sets of analyses: (1) We used ordered logit regressions to estimate the direct relationships of the supply chain variables, future partner plans, and the outcomes, and (2) we used ordered logit regressions to estimate the mediation effects of the supply chain partner choices on outcomes. The variables were ordinal ones to drive the decision for ordered logit regressions. This approach does not assume equal intervals between levels in the dependent variable. The ordered logit model is as follows:

$$Y^*_i = \beta X_i + e_i,$$

where Y^*_i is the propensity of respondents to indicate higher levels of the dependent variables (ie, SC-DISRUP, GROWTH, SERVICE-IMPR, SC-INTEGR, SC-REDEGN), X_i is a set of explanatory variables, β is a vector of parameters, and e_i are disturbances.

We did not observe Y^*_i. Instead, we observed the ordinal dependent variable Y_i depending on the values of thresholds or cut-off points τ_{m-1} and τ_m, and the probability distribution of Y_i is given as follows:

$$Pr = [Y_i = m|X_i = F(\tau_m - X\beta) - F(\tau_{m-1} - X\beta)]$$

Results

Sample Statistics

The descriptive statistics and pairwise correlations among the key variables used in this study are presented in Tables 3 and 4. As shown in Table 3, health systems face relatively high supply chain disruptions and challenges (SC-DISRUP mean

4.45, SD 1.82). Additionally, the anticipation of improving health delivery and services overall (SERVICE-IMPR mean 3.48, SD 2.07) seemed to be a more popular outlook among CEOs over the presentation of growth opportunities (GROWTH mean 3.08, SD 2.01). The partnership choice of integrating with supply chain and logistics organizations (SC-INTEGR mean 3.98, SD 2.15) is similar to redesign through start-ups or entrepreneurial collaborations (SC-REDEGN mean 3.97, SD 2.15). Table 4 presents the correlations between the main variables.

Table 3. Summary statistics (N=124).

Variable	Mean (SD)	Min.	Max.
SC-DISRUP[a]	4.45 (1.82)	1	7
GROWTH[b]	3.08 (2.01)	1	7
SERVICE-IMPR[c]	3.48 (2.07)	1	7
SC-INTEGR[d]	3.98 (2.15)	1	7
SC-REDEGN[e]	3.97 (2.15)	1	7
SIZE_B-SMALL	0.08 (0.27)	0	1
SIZE_B-MEDIUM	0.40 (0.49)	0	1
SIZE_B-LARGE	0.52 (0.50)	0	1
REGION-NE[f]	0.22 (0.42)	0	1
REGION-MW[g]	0.26 (0.44)	0	1
REGION-SOUTH	0.34 (0.47)	0	1
REGION-WEST	0.18 (0.38)	0	1
TEACHING-NON	0.30 (0.46)	0	1
TEACHING-MINOR	0.44 (0.50)	0	1
TEACHING-MAJOR	0.26 (0.44)	0	1
REVENUE-LOW	0.12 (0.33)	0	1
REVENUE-MEDIUM	0.05 (0.22)	0	1
REVENUE-HIGH	0.83 (0.38)	0	1
HIGH-DSH[h]-HOSP	0.33 (0.47)	0	1
HIGH-BURDEN-SYS	0.19 (0.39)	0	1
HIGH-BURDEN-HOSP	0.32 (0.47)	0	1
OWNERSHIP	0.03 (0.16)	0	1
PHYSICIANS	1.86 (0.77)	1	3
HOSPITALS	1.50 (0.81)	1	3

[a]SC-DISRUP: supply chain disruptions and challenges.

[b]GROWTH: presenting growth opportunities in the post-COVID-19 new normal.

[c]SERVICE-IMPR: overall improvements in health delivery and services.

[d]SC-INTEGR: integration with supply chain and logistics organizations.

[e]SC-REDEGN: redesign through start-ups or entrepreneurial collaborations.

[f]NE: Northeast.

[g]MW: Midwest.

[h]DSH: disproportionate share hospital.

Table 4. Pairwise correlations between main variables (N=124).

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	SC-DISRUP[a]	1.00	-.03	.25	.69	-.46	.02	-.04	.14	-.23	-.01	.07	-.06	.08	-.13	.06	-.11	.04	-.08	-.01	.02
2	GROWTH[b]	–.03	1.00	.26	.07	.17	-.14	.22	-.06	.04	-.01	-.10	.03	.03	-.01	.04	-.01	.06	-.12	.16	.13
3	SERVICE-IMPR[c]	.25	.26	1.00	.20	-.04	.04	.06	.10	-.07	.03	-.12	-.03	.08	-.11	.10	-.06	.13	.07	.10	.02
4	SC-INTEGR[d]	.69	.07	.20	1.00	-.23	.03	-.08	.03	-.11	-.01	.04	-.04	-.22	-.25	-.08	-.07	.10	-.11	-.15	-.05
5	SC-REDEGN[e]	–.46	.17	–.04	–.23	1.00	-.05	.09	-.02	.17	.12	-.15	-.06	.08	.15	-.17	.05	-.05	.01	-.01	.00
6	SIZE_B-MED	.02	–.14	.04	.03	–.05	1.00	-.83	-.08	-.01	-.05	.03	-.10	-.16	-.27	-.32	-.08	-.01	-.21	-.60	-.46
7	SIZE_B-LARGE	–.04	.22	.06	–.08	.09	–.83	1.00	-.01	.11	.01	-.14	.26	.25	.36	.38	.17	.06	.28	.75	.54
8	REGION-MW[f]	.14	–.06	.10	.03	–.02	–.08	–.01	1.00	-.42	-.26	-.07	-.01	-.12	-.09	.09	-.20	-.10	.00	.02	.02
9	REGION-SOUTH	–.23	.04	–.07	–.11	.17	–.01	.11	–.42	1.00	-.34	.17	.07	.02	.15	-.02	.12	.17	.22	.00	.04
10	REGION-WEST	–.01	–.01	.03	–.01	.12	–.05	.005	–.26	–.34	1.00	-.06	-.02	-.04	-.06	-.01	.12	.08	.02	-.04	-.03
11	TEACHING-MINOR	.07	–.10	–.12	.04	–.15	.03	–.14	–.07	.17	–.06	1.00	.01	-.07	-.07	-.05	.05	-.06	-.08	-.05	-.08
12	TEACHING-MAJOR	–.06	.03	–.03	–.04	–.06	–.10	.26	–.01	.07	–.02	.01	1.00	-.50	.12	.07	.05	-.12	.08	.19	.26
13	REVENUE-MEDIUM	–.08	.03	.08	–.22	.08	–.16	.25	–.12	.02	–.04	–.07	–.50	1.00	.12	.17	.34	.03	.13	.38	.06
14	REVENUE-HIGH	–.13	–.01	–.11	–.25	.15	–.27	.36	–.09	.15	–.06	–.07	.12	.12	1.00	-.23	-.06	-.04	.14	.26	.29
15	HIGH-DSH[g]-HOSP	.06	.04	.10	–.08	–.17	–.32	.38	.09	–.02	–.01	–.05	.07	.17	–.23	1.00	.15	-.04	-.02	.51	.30
16	HIGH-BURDEN-SYS	–.11	–.01	–.06	–.07	.05	–.08	.17	–.20	.12	.12	.05	.05	.34	–.06	.15	1.00	-.01	.18	.23	.17
17	HIGH-BURDEN-HOSP	.04	.06	.13	.10	–.05	–.01	.06	–.10	.17	.08	–.06	–.12	.03	–.04	–.04	–.01	1.00	.42	-.10	-.20
18	OWNERSHIP	–.08	–.12	.07	–.11	.01	–.21	.28	.002	.22	.02	–.08	.08	.13	.14	–.02	.18	.42	1.00	.18	.31
19	PHYSICIANS	–.01	.16	.10	–.15	–.01	–.60	.75	.02	.002	–.04	–.05	.19	.38	.26	.51	.23	–.10	.18	1.00	.57
20	HOSPITALS	.02	.13	.02	–.05	.001	–.46	.54	.02	.04	–.03	–.08	.26	.06	.29	.30	.17	–.20	.31	.57	1.00

[a]SC-DISRUP: supply chain disruptions and challenges.

[b]GROWTH: presenting growth opportunities in the post-COVID-19 new normal.

[c]SERVICE-IMPR: overall improvements in health delivery and services.

[d]SC-INTEGR: integration with supply chain and logistics organizations.

[e]SC-REDEGN: redesign through start-ups or entrepreneurial collaborations.

[f]MW: Midwest.

[g]DSH: disproportionate share hospital.

Estimation Outcomes

Table 5 shows the first set of results of the ordered logit model estimation that describes the relationship between perception of supply chain–related challenges (SC-DISRUP) and future partnership plans: either integrating with supply chain and logistics organizations (SC-INTEGR) or redesigning through start-ups or entrepreneurial collaborations (SC-REDEGN).

We found a positive and significant relationship between the supply chain issue and supply chain partner choice. With a higher expectation of supply chain disruption and challenges, health systems partner with supply chain and logistics organizations and integrate it better. In addition, there was a negative and significant relationship between supply chain issue and start-up partner choice, indicating that with higher expectations of supply chain disruption and challenges, health systems tend *not* to partner with start-ups or entrepreneurial collaborations (ie, redesigning supply chain using new start-ups).

Table 6 shows the direct relationships between supply chain partnership plans and outcomes. First, there are positive and significant relationships between supply chain partner and service in the separate models and between start-up partner and growth. Both partner options can contribute to the growth in the combined models, while supply chain partners can significantly increase service improvement. In other words, integration is better for service and growth, but redesign is better for growth. The results indicate that with solid supply chain support by integration, health systems can provide better services; with redesign, health systems can increase the innovation/capability needed for growth.

We coded a new variable to capture such dual options to explore the impacts of supply chain partnership plans when health systems have integration and redesign plans. This variable is a dummy variable that equals 1 if both SC-INTEGR and SC-REDEGN are greater than their mean values, or 0 otherwise. Table 7 shows the direct relationships between supply chain

challenges and dual-partnership (DUAL) plans and between dual-partnership plans and outcomes. We found no significant relationship between perceived supply chain challenges and dual-partnership choices. There are significant relationships between dual-partnership choices and growth and service outcomes, indicating that health systems can increase growth and service improvement when health systems choose both partner options.

After examining the direct relationships, we tested the mediating effects of supply chain partnership plans on the outcomes. Table 8 displays the impacts on growth, and Table 9 presents the impacts on service. We found that supply chain challenges do not directly affect growth and service outcomes, and partnership plans can mediate these relationships. We further conducted the Sobel mediation test to check the mediation effects. We found that supply chain integration partner choice has a significant and positive impact on growth outcome; the mediating effect is 0.84 (proportion of the total effect that is mediated). In addition, the redesign through start-up choice has a significant and positive impact on growth outcome (DV), and the mediating effect is 0.75. In contrast, for the impacts on service outcome, we found that the mediating effect of supply chain integration partner choice is only 0.18, while the mediating effect of redesign through start-up choice is only 0.08.

Table 5. Direct relationship between supply chain perceived challenges and partnership plans.[a]

Variables	SC-INTEGR[b,c]		SC-REDEGN[d,e]	
	Coefficient (SE)	*P* value	Coefficient (SE)	*P* value
SC-DISRUP[f]	.858 (0.049)	<.001	–.478 (0.058)	<.001
SIZE	.599 (0.486)	.22	.693 (0.060)	<.001
REGION	–.129 (0.217)	.55	.466 (0.133)	<.001
OWNERSHIP	–.522 (0.598)	.38	–1.541 (0.534)	.004
TEACHING	–.962 (0.110)	<.001	–.122 (0.490)	.80
REVENUE	–.826 (0.270)	.002	–.479 (0.494)	.33
HIGH-DSH[g]-HOSP	.839 (0.489)	.09	–.025 (0.176)	.89
HIGH-BURDEN-SYS	.400 (0.301)	.18	–.547 (0.069)	<.001
HIGH-BURDEN-HOSP	–.257 (0.092)	.01	–.374 (0.231)	.11
PHYSICIANS	–.016 (0.444)	.97	–.082 (0.112)	.47
HOSPITALS	.268 (0.144)	.06	.018 (0.182)	.92

[a]The results of the cut-off points are omitted for brevity.

[b]SC-INTEGR: integration with supply chain and logistics organizations.

[c]Pseudo R^2=.2038 (n=123 observations).

[d]SC-REDEGN: redesign through start-ups or entrepreneurial collaborations.

[e]Pseudo R^2=.1112 (n=124 observations).

[f]SC-DISRUP: supply chain disruptions and challenges.

[g]DSH: disproportionate share hospital.

Table 6. Direct relationship between supply chain partnership plans and outcomes.[a]

Variables	GROWTH[b,c]		SERVICE-IMPR[d,e]		GROWTH[f]		SERVICE-IMPR[g]		GROWTH[h]		SERVICE-IMPR[i]	
	Coefficient (SE)	*P* value	Coefficient (SE)	*P* value	Coefficient (SE)	*P* value	Coefficient (SE)	*P* value	Coefficient (SE)	*P* value	Coefficient (SE)	*P* value
SC-INTEGR[j]	.096 (0.060)	.11	.327 (0.117)	.005	N/A[k]	N/A	N/A	N/A	.142 (0.038)	*P*<.001	.335 (0.120)	.01
SC-REDEGN[l]	N/A	N/A	N/A	N/A	.200 (0.086)	.02	–.010 (0.125)	.94	.212 (0.083)	.01	.037 (0.140)	.79
SIZE	.915 (0.362)	.01	.165 (0.597)	.78	.852 (0.380)	.03	.319 (0.472)	.50	.793 (0.377)	.04	.145 (0.512)	.78
REGION	.053 (0.173)	.76	.182 (0.066)	.006	–.116 (0.207)	.58	.078 (0.102)	.45	–.076 (0.256)	.77	.165 (0.121)	.17
OWNERSHIP	–.792 (0.922)	.39	–1.484 (0.766)	.05	–.221 (0.642)	.73	–1.298 (0.377)	.001	–.210 (0.724)	.77	–1.378 (0.490)	.01
TEACHING	–.187 (0.133)	.16	.295 (0.036)	<.001	–.199 (0.030)	<.001	.082 (0.062)	.19	–.151 (0.008)	<.001	.293 (0.017)	<.001
REVENUE	–.762 (0.199)	<.001	–.133 (0.237)	.58	–.727 (0.246)	.003	–.320 (0.176)	.07	–.622 (0.265)	.02	–.117 (0.254)	.65
HIGH-DSH[m]-HOSP	–.207 (0.174)	.23	–.461 (0.266)	.08	–.124 (0.072)	.09	–.345 (0.173)	.05	–.204 (0.137)	.14	–.460 (0.261)	.08
HIGH-BURDEN-SYS	.867 (0.289)	.003	.313 (0.273)	.25	1.119 (0.396)	.01	.525 (0.142)	<.001	1.055 (0.305)	.001	.326 (0.239)	.17
HIGH-BURDEN-HOSP	–1.363 (0.229)	<.001	–.098 (0.169)	.56	–1.457 (0.116)	<.001	–.172 (0.100)	.09	–1.415 (0.142)	<.001	–.092 (0.146)	.53
PHYSICIANS	.312 (0.080)	<.001	.270 (0.218)	.22	.291 (0.125)	.02	.257 (0.157)	.10	.314 (0.165)	.06	.276 (0.196)	.16
HOSPITALS	.442 (0.055)	<.001	–.149 (0.096)	.12	.532 (0.039)	<.001	–.035 (0.100)	.73	.460 (0.067)	<.001	–.146 (0.085)	.08

[a]The results of the cut-off points are omitted for brevity.

[b]GROWTH: presenting growth opportunities in the post-COVID-19 new normal.

[c]Pseudo R^2=.0583 (N= 123 observations).

[d]SERVICE-IMPR: overall improvements in health delivery and services.

[e]Pseudo R^2=.0347 (N= 123 observations).

[f]Pseudo R^2=.0689 (N= 124 observations).

[g]Pseudo R^2=.0162 (N= 124 observations).

[h]Pseudo R^2=.0717 (N= 123 observations).

[i]Pseudo R^2=.0350 (N= 123 observations).

[j]SC-INTEGR: integration with supply chain and logistics organizations.

[k]N/A: not applicable.

[l]SC-REDEGN: redesign through start-ups or entrepreneurial collaborations.

[m]DSH: disproportionate share hospital.

Table 7. Direct relationship between supply chain challenges, dual-partnership plans, and outcomes.[a]

Variables	DUAL[b,c]		GROWTH[d,e]		SERVICE-IMPR[f,g]	
	Coefficient (SE)	*P* value	Coefficient (SE)	*P* value	Coefficient (SE)	*P* value
SC-DISRUP[h]	.035 (0.135)	.796	N/A[i]	<.001	N/A	N/A
DUAL	N/A	N/A	.848 (0.053)	<.001	.696 (0.354)	.05
SIZE	.526 (0.169)	.002	.916 (0.299)	.002	.254 (0.408)	.53
REGION	.293 (0.302)	.33	–.058 (0.184)	.75	.056 (0.164)	.74
OWNERSHIP	–14.273 (1.144)	<.001	–.298 (0.747)	.69	–.920 (1.297)	.48
TEACHING	–.294 (0.763)	.70	–.167 (0.011)	<.001	.133 (0.308)	.67
REVENUE	–1.150 (0.095)	<.001	–.648 (0.201)	.001	–.188 (0.305)	.54
HIGH-DSH[j]-HOSP	–.179 (0.384)	.64	–.069 (0.146)	.64	–.323 (0.375)	.39
HIGH-BURDEN-SYS	.590 (0.220)	.01	.921 (0.365)	.01	.429 (0.510)	.40
HIGH-BURDEN-HOSP	–.633 (0.149)	<.001	–1.408 (0.185)		–.145 (0.443)	.74
PHYSICIANS	.370 (0.572)	.52	.158 (0.117)	.18	.210 (0.342)	.54
HOSPITALS	.247 (0.151)	.10	.537 (0.066)	<.001	–.051 (0.276)	.85

[a]The results of the cut-off points are omitted for brevity.

[b]DUAL: dual partnership.

[c]Pseudo R^2=.1067 (N=124 observations).

[d]GROWTH: presenting growth opportunities in the post-COVID-19 new normal.

[e]Pseudo R^2=.0703 (N=124 observations).

[f]SERVICE-IMPR: overall improvements in health delivery and services.

[g]Pseudo R^2=.0257 (N=124 observations).

[h]SC-DISRUP: supply chain disruptions and challenges.

[i]N/A: not applicable.

[j]DSH: disproportionate share hospital.

Table 8. Mediating effects of supply chain partnership plans on growth outcomes.[a]

Variables	GROWTH[b,c]		GROWTH[d]		GROWTH[e]		GROWTH[f]	
	Coefficient (SE)	*P* value	Coefficient (SE)	*P* value	Coefficient (SE)	*P* value	Coefficient (SE)	*P* value
SC-DISRUP[g]	–.064 (0.053)	.23	–.170 (0.073)	.02	.028 (0.023)	.22	–.068 (0.061)	.27
SC-INTEGR[h]	N/A[i]	N/A	.224 (0.048)	<.001	N/A	N/A	.189 (0.065)	.004
SC-REDEGN[j]	N/A	N/A	N/A	N/A	.209 (0.085)	.01	.195 (0.085)	.02
SIZE	.954 (0.374)	.01	.873 (0.391)	.03	.848 (0.372)	.02	.787 (0.389)	.04
REGION	–.006 (0.134)	.96	.028 (0.179)	.88	–.116 (0.210)	.58	–.073 (0.256)	.78
OWNERSHIP	–.702 (0.787)	.37	–.616 (0.883)	.49	–.226 (0.640)	.72	–.186 (0.725)	.80
TEACHING	–.230 (0.129)	.08	–.181 (0.116)	.12	–.187 (0.018)	<.001	–.153 (0.011)	<.001
REVENUE	–.846 (0.172)	<.001	–.717 (0.163)	<.001	–.720 (0.249)	.004	–.614 (0.249)	.01
HIGH-DSH[k]-HOSP	–.135 (0.107)	.21	–.250 (0.175)	.15	–.123 (0.075)	.10	–.224 (0.153)	.14
HIGH-BURDEN-SYS	.950 (0.405)	.02	.889 (0.324)	.01	1.114 (0.384)	.004	1.045 (0.312)	.001
HIGH-BURDEN-HOSP	–1.394 (0.165)	<.001	–1.340 (0.186)	<.001	–1.454 (0.124)	<.001	–1.400 (0.141)	<.001
PHYSICIANS	.301 (0.064)	<.001	.354 (0.124)	.004	.280 (0.127)	.03	.332 (0.192)	.09
HOSPITALS	.522 (0.062)	<.001	.465 (0.051)	<.001	.527 (0.041)	<.001	.467(0.046)	<.001

[a]The results of the cut-off points are omitted for brevity.

[b]GROWTH: presenting growth opportunities in the post-COVID-19 new normal.

[c]Pseudo R^2=.0578 (N=124 observations).

[d]Pseudo R^2=.0625 (N=123 observations).

[e]Pseudo R^2=.0691 (N=124 observations).

[f]Pseudo R^2=.0723 (N=123 observations).

[g]SC-DISRUP: supply chain disruptions and challenges.

[h]SC-INTEGR: integration with supply chain and logistics organizations.

[i]N/A: not applicable.

[j]SC-REDEGN: redesign through start-ups or entrepreneurial collaborations.

[k]DSH: disproportionate share hospital.

Table 9. Mediating effects of supply chain partnership plans on service outcomes.[a]

Variables	SERVICE-IMPR[b,c]		SERVICE-IMPR[d]		SERVICE-IMPR[e]		SERVICE-IMPR[f]	
	Coefficient (SE)	*P* value	Coefficient (SE)	*P* value	Coefficient (SE)	*P* value	Coefficient (SE)	*P* value
SC-DISRUP[g]	.289 (0.113)	.01	.201 (0.101)	.05	.310 (0.089)	.001	.221 (0.060)	<.001
SC-INTEGR[h]	N/A[i]	N/A	.178 (0.130)	.17	N/A	N/A	.177 (0.133)	.18
SC-REDEGN[j]	N/A	N/A	N/A	N/A	.073 (0.135)	.59	.071 (0.136)	.60
SIZE	.303 (0.544)	.58	.216 (0.563)	.70	.261 (0.454)	.57	.179 (0.478)	.71
REGION	.143 (0.047)	.002	.197 (0.059)	.001	.109 (0.094)	.25	.164 (0.119)	.17
OWNERSHIP	–1.680 (0.637)	.01	–1.684 (0.690)	.02	–1.489 (0.461)	.001	–1.498 (0.499)	.003
TEACHING	.221 (0.041)	<.001	.274 (0.034)	<.001	.219 (0.026)	<.001	.271 (0.017)	<.001
REVENUE	–.293 (0.262)	.26	–.179 (0.292)	.54	–.265 (0.284)	.35	–.153 (0.323)	.64
HIGH-DSH[k]-HOSP	–.331 (0.195)	.09	–.416 (0.259)	.11	–.330 (0.193)	.09	–.414 (0.255)	.10
HIGH-BURDEN-SYS	.321 (0.258)	.21	.268 (0.328)	.41	.353 (0.233)	.13	.295 (0.315)	.35
HIGH-BURDEN-HOSP	–.038 (0.236)	.87	–.056 (0.237)	.81	–.025 (0.207)	.90	–.042 (0.209)	.84
PHYSICIANS	.185 (0.221)	.40	.240 (0.206)	.24	.194 (0.196)	.32	.248 (0.186)	.18
HOSPITALS	–.120 (0.109)	.27	–.188 (0.112)	.09	–.115 (0.107)	.28	–.185 (0.115)	.11

[a]The results of the cut-off points are omitted for brevity.

[b]SERVICE-IMPR: overall improvements in health delivery and services.

[c]Pseudo R^2=.0353 (N=124 observations).

[d]Pseudo R^2=.0403 (N=123 observations).

[e]Pseudo R^2=.0365 (N=124 observations).

[f]Pseudo R^2=.0414 (N=123 observations).

[g]SC-DISRUP: supply chain disruptions and challenges.

[h]SC-INTEGR: integration with supply chain and logistics organizations.

[i]N/A: not applicable.

[j]SC-REDEGN: redesign through start-ups or entrepreneurial collaborations.

[k]DSH: disproportionate share hospital.

Discussion

Principal Findings

This study first explores the relationship between the perceived severity of supply chain–related challenges/disruptions and an organization's future partnership plans to address those challenges. Second, the study elucidates the relationship between supply chain partnerships and outcomes. We also explored the relationship between supply chain challenges and a dual-partnership mix.

First, the results indicate that organizations tend to choose supply chain integration with existing partners as an operational strategy with higher perceived challenges and disruptions. These organizations tend not to partner with start-ups or other new entities.

Health systems with HIGH-DSH-HOSP and HIGH-BURDEN-SYS both tend to choose integration over collaboration. REGION and REVENUE negatively correlate with integration and positively correlate with collaboration. OWNERSHIP, TEACHING hospitals, HIGH-BURDEN-HOSP, and PHYSICIANS negatively correlate with integration and collaboration. SIZE and HOSPITALS both have a positive correlation with integration and collaboration.

Second, the results indicate a significant relationship between supply chain integration and service improvement and between supply chain collaborations and growth. Together, integration and collaboration contribute to growth, while integration has a more profound positive impact on service improvement. Thus, integration is better for both service and growth, while collaboration is better for growth.

There is a positive correlation with both GROWTH and SERVICE-IMPR for SIZE, REGION, HIGH-BURDEN-SYS, and PHYSICIANS in health systems that choose to integrate with existing partners. OWNERSHIP, REVENUE, HIGH-DSH-HOSP, and HIGH-BURDEN-HOSP negatively correlate with GROWTH and SERVICE-IMPR. TEACHING hospitals negatively correlate with GROWTH and have a positive correlation with SERVICE-IMPR. Thus, hospitals have a positive correlation with GROWTH and a negative correlation with SERVICE-IMPR.

For health systems that choose supply chain redesign, there is a positive correlation with both GROWTH and SERVICE-IMPR for the variables SIZE, HIGH-BURDEN-SYS, and

PHYSICIANS. There is a negative correlation with GROWTH and a positive correlation with SERVICE-IMPR for REGION and TEACHING. There is a negative correlation with both GROWTH and SERVICE-IMPR for OWNERSHIP, REVENUE, HIGH-DSH-HOSP, and HIGH-BURDEN-HOSP. Thus, hospitals have a positive correlation with GROWTH and a negative correlation with SERVICE-IMPR.

Third, the results indicate no significant relationship between supply chain challenges and dual-partnership choice; however, dual-partnership choices significantly affected both service and growth. Taken together, the results suggest that health systems are better able to provide services with supply chain integration. Health systems positioned for growth can also increase innovation and capability through supply chain redesign.

For health systems that have dual partnerships, there is a positive correlation with both GROWTH and SERVICE-IMPR for the variables SIZE, HIGH-BURDEN-SYS, and PHYSICIANS. There is a negative correlation with GROWTH and a positive correlation with SERVICE-IMPR for REGION and TEACHING. There is a negative correlation with GROWTH and SERVICE-IMPR for OWNERSHIP, REVENUE, HIGH-DSH-HOSP, and HIGH-BURDEN-HOSP. Thus, hospitals have a positive correlation with GROWTH and a negative correlation with SERVICE-IMPR.

Implications

The first set of results sheds light on the influence of perceived disruptions and challenges on partnership choice. Higher expectations of supply chain disruptions tend to influence supply chains to integrate more with existing partners and not with start-ups or other entrepreneurial collaborations. A plausible explanation is that high-burden systems may not want to pursue new collaborations to further add to the complexity of their current state. Additionally, health systems with high revenue may opt for supply chain redesign because of the financial resources available for the collaborative initiative.

The second set of findings investigates partnership choice and outcomes. Supply chain integration leads to greater service outcomes, while collaborations lead to greater growth opportunities. Although integration shows a more profound impact on service improvement, there are growth opportunities within this partnership choice. One possible explanation for these findings is that large high-burden systems have greater opportunity for growth within the supply chain when operations become more integrated and the flow of resources becomes more seamless. Resources can be redirected to growth initiatives.

Examining the factors that lead organizations to integration or collaboration will guide new strategies and policies in health care. Policy implementations that can minimize the financial risks of collaborations may drive dual-partnership choices. Healthy systems and health systems that do not perceive many challenges have an opportunity for competitive advantage through start-up collaborations. Additionally, organizations may contract with alternative suppliers for emergencies to ensure supply procurement in events that halt shipments. Future research could elucidate the impact of highly networked systems on perceived supply chain disruptions and thus partnership choice.

Limitations and Directions for Future Research

The authors acknowledge some limitations of this study that future research could address. Integration and collaboration may not lead to increased service outcomes to growth opportunities if they are not implemented successfully, so providers, members of the supply chain team, and policymakers should focus on the conditions with which integration and collaboration can be successfully implemented to deliver greater value in the care continuum.

First, research could further examine other variables affecting partnership choice. This study focused on the perception of supply chain disruptions and challenges. However, other external and internal disruptions could affect partnership choice. Demand-driven externalities could also influence partnership choice.

Second, supply chain operations can be highly networked or have few key players. In health systems with few suppliers, an impact on a critical supplier will cause more significant disruptions in production. Organizations may collaborate with other suppliers in this case because further integration in the supply chain may not increase supply procurement.

Conclusion

Disruptions in health care remain a challenge, partly due to supply chain management inefficiencies. Understanding these inefficiencies and how they impact supply chain partnership choice as an organizational strategy is essential in the health care sector. This study begins to investigate supply chain management partnership dynamics.

Health care organizations continuously face growing costs and meet these challenges through seamless integration efforts and innovative collaborations. Although higher perceptions of supply chain challenges lead to integration strategies rather than innovative redesign, integration and collaboration can provide lean and agile systems and better equip organizations to address challenges. Organizations have an opportunity to provide better service outcomes while also identifying and capitalizing on innovative growth opportunities.

Acknowledgments

This research is part of the Health Systems' Climate Study of 2021 conducted by the Health Administration Research Consortium. The Climate Study aims to understand the current state of health systems in the United States following the COVID-19 pandemic. Our sincere thanks to the active and candid participation of the chief executive officers (CEOs) from 135 health systems. The authors thank Naser Shekarian, PhD candidate in the Computer Science and Information Systems (CSIS) Business PhD program at the University of Colorado Denver, for help in compiling the CEOs' addresses for the climate study data collection. The authors

also thank the support from the Business School at the University of Colorado Denver for this project, specifically the Health Administrations and the CSIS Business PhD programs for the time and efforts of the authors involved with this study.

Conflicts of Interest

None declared.

References

1. Mazareanu E. Impact of the Coronavirus Pandemic on Supply Chains across Industries 2020. URL: https://www.statista.com/statistics/1182233/impact-covid-19-supply-chain/ [accessed 2022-04-29]
2. Guest JL, Del Rio C, Sanchez T. The three steps needed to end the COVID-19 pandemic: bold public health leadership, rapid innovations, and courageous political will. JMIR Public Health Surveill 2020 Apr 06;6(2):e19043 [FREE Full text] [doi: 10.2196/19043] [Medline: 32240972]
3. Khuntia J, Ning X, Stacey R. Competition and integration of US health systems in the post-COVID-19 new normal: cross-sectional survey. JMIR Form Res 2022 Mar 24;6(3):e32477 [FREE Full text] [doi: 10.2196/32477] [Medline: 35133973]
4. Kwon IG, Kim S, Martin DG. Healthcare supply chain management; strategic areas for quality and financial improvement. Technol Forecast Soc Change 2016 Dec;113:422-428. [doi: 10.1016/j.techfore.2016.07.014]
5. Mathew J, John J, Kumar S, Management O. New trends in healthcare supply chain. 2013 Presented at: POM 24th Annual Conference of the Production and Operations Management; 2013; Denver, CO p. 1-10.
6. Bhakoo V, Singh P, Sohal A. Collaborative management of inventory in Australian hospital supply chains: practices and issues. Supply Chain Manag 2012;17:217-230. [doi: 10.1108/13598541211212933]
7. Aronsson H, Abrahamsson M, Spens K. Developing lean and agile health care supply chains. Supply Chain Manag 2011 May 03;16(3):176-183. [doi: 10.1108/13598541111127164]
8. Francis JR. COVID-19: implications for supply chain management. Front Health Serv Manag 2020;37(1):33-38. [doi: 10.1097/HAP.0000000000000092] [Medline: 32842087]
9. Iyengar KP, Vaishya R, Bahl S, Vaish A. Impact of the coronavirus pandemic on the supply chain in healthcare. Br J Healthc Manag 2020 Jun 02;26(6):1-4. [doi: 10.12968/bjhc.2020.0047]
10. Nikolopoulos K, Punia S, Schäfers A, Tsinopoulos C, Vasilakis C. Forecasting and planning during a pandemic: COVID-19 growth rates, supply chain disruptions, and governmental decisions. Eur J Oper Res 2021 Apr 01;290(1):99-115 [FREE Full text] [doi: 10.1016/j.ejor.2020.08.001] [Medline: 32836717]
11. Falasca M, Zobel C, Cook D. A decision support framework to assess supply chain resilience. 2008 Presented at: Proceedings of ISCRAM 2008 - 5th International Conference on Information Systems for Crisis Response and Management; 2008; Washington, DC p. 596-605.
12. Dun & Bradstreet. Business Impact of the Coronavirus: Business and Supply Chain Analysis Due to the Coronavirus Outbreak. URL: https://dnb.com/content/dam/english/economic-and-industry-insight/DNB_Business_Impact_of_the_Coronavirus_US.pdf [accessed 2022-04-29]
13. Gunasekaran A, Ngai E. Information systems in supply chain integration and management. Eur J Oper Res 2004 Dec;159(2):269-295. [doi: 10.1016/j.ejor.2003.08.016]
14. Kim SW, Narasimhan R. Information system utilization in supply chain integration efforts. Int J Prod Res 2002 Jan;40(18):4585-4609. [doi: 10.1080/0020754021000022203]
15. Gonul Kochan C, Nowicki DR, Sauser B, Randall WS. Impact of cloud-based information sharing on hospital supply chain performance: a system dynamics framework. Int J Prod Econ 2018 Jan;195:168-185. [doi: 10.1016/j.ijpe.2017.10.008]
16. Wang W, Chen L, Zhang Q. Outsourcing high-dimensional healthcare data to cloud with personalized privacy preservation. Comput Netw 2015 Sep;88:136-148. [doi: 10.1016/j.comnet.2015.06.014]
17. Barratt M. Understanding the meaning of collaboration in the supply chain. Supply Chain Manag 2004 Feb;9(1):30-42. [doi: 10.1108/13598540410517566]
18. Khuntia J, Ning X, Stacey R. Digital orientation of health systems in the post-COVID-19 "new normal" in the United States: cross-sectional survey. J Med Internet Res 2021 Aug 16;23(8):e30453 [FREE Full text] [doi: 10.2196/30453] [Medline: 34254947]
19. Kampstra R, Ashayeri J, Gattorna J. Realities of supply chain collaboration. Int J Logist Manag 2006 Sep;17(3):312-330. [doi: 10.1108/09574090610717509]
20. Cao M, Vonderembse MA, Zhang Q, Ragu-Nathan T. Supply chain collaboration: conceptualisation and instrument development. Int J Prod Res 2009 Dec 14;48(22):6613-6635. [doi: 10.1080/00207540903349039]
21. Gaynor M, Tuttle-Newhall J, Parker J, Patel A, Tang C. Adoption of blockchain in health care. J Med Internet Res 2020 Sep 17;22(9):e17423 [FREE Full text] [doi: 10.2196/17423] [Medline: 32940618]
22. Khurshid A. Applying blockchain technology to address the crisis of trust during the COVID-19 pandemic. JMIR Med Inform 2020 Sep 22;8(9):e20477 [FREE Full text] [doi: 10.2196/20477] [Medline: 32903197]

23. Cichosz M. Logistics startups and logistics service providers: competitors or partners in exploration? In: Suchanek M, editor. Transport Development Challenges in the 21st Century. TranSopot 2019. Springer Proceedings in Business and Economics. Cham: Springer; 2021:1-11.
24. Ishii L, Demski R, Ken Lee KH, Mustafa Z, Frank S, Wolisnky JP, et al. Improving healthcare value through clinical community and supply chain collaboration. Healthc (Amst) 2017 Mar;5(1-2):1-5. [doi: 10.1016/j.hjdsi.2016.03.003] [Medline: 28668197]
25. Langabeer J, Helton J. Health Care Operations Management. 3rd ed. Burlington, MA: Jones & Bartlett; 2021.
26. Hut N. Reimagining the Healthcare Supply Chain to Bolster Resilience and Efficiency. URL: https://www.hfma.org/content/hfma/www/topics/hfm/2020/november/reimagining-the-healthcare-supply-chain-to-bolster-resilience-an/ [accessed 2022-04-29]
27. Infosys BPM. How Healthcare Supply Chains Can Move towards post-COVID Resilience. URL: https://www.infosysbpm.com/blogs/sourcing-procurement/building-a-resilient-healthcare-supply-chain-post-covid.html [accessed 2022-04-29]
28. Motiwala SS, McLaughlin JE, King J, Hodgson B, Hamilton M. Advancing the health care supply chain and promoting leadership through strategic partnerships with industry. Healthc Manag Forum 2008 Jun 01;21(2):23-28. [doi: 10.1016/s0840-4704(10)60542-x]
29. Nandi S, Sarkis J, Hervani AA, Helms MM. Redesigning supply chains using blockchain-enabled circular economy and COVID-19 experiences. Sustain Prod Consum 2021 Jul;27:10-22 [FREE Full text] [doi: 10.1016/j.spc.2020.10.019] [Medline: 33102671]
30. Khuntia J, Ning X, Shekarian N, Stacey R. Inaugural Health Systems' Climate Study 2021. Denver, CO: University of Colorado Business School; 2021.

Abbreviations

AHRQ: Agency for Healthcare Research and Quality
CEO: chief executive officer
DSH: disproportionate share hospital
DUAL: dual partnership
GPO: group purchasing organization
GROWTH: presenting growth opportunities in the post-COVID-19 new normal
MW: Midwest
NE: Northeast
PPE: personal protective equipment
SC-DISRUP: supply chain disruptions and challenges
SC-INTEGR: integration with supply chain and logistics organizations
SC-REDEGN: redesign through start-ups or entrepreneurial collaborations
SERVICE-IMPR: overall improvements in health delivery and services

Edited by T Leung; submitted 30.11.21; peer-reviewed by J Langabeer, Y Tani; comments to author 01.03.22; revised version received 09.04.22; accepted 21.04.22; published 15.06.22.

Please cite as:
Khuntia J, Mejia FJ, Ning X, Helton J, Stacey R
Integration vs Collaborative Redesign Strategies of Health Systems' Supply Chains in the Post-COVID-19 New Normal: Cross-sectional Survey Across the United States
JMIR Form Res 2022;6(6):e35317
URL: https://formative.jmir.org/2022/6/e35317
doi:10.2196/35317
PMID:35452405

Original Paper

A Scalable Risk-Scoring System Based on Consumer-Grade Wearables for Inpatients With COVID-19: Statistical Analysis and Model Development

Simon Föll[1*], MSc; Adrian Lison[1*], MSc; Martin Maritsch[1*], MSc; Karsten Klingberg[2*], MD; Vera Lehmann[3], MD; Thomas Züger[1,3,4], MD; David Srivastava[2], MD; Sabrina Jegerlehner[2], MD; Stefan Feuerriegel[1,5], PhD; Elgar Fleisch[1,6], PhD; Aristomenis Exadaktylos[2], MD; Felix Wortmann[1,6], PhD

[1]Department of Management, Technology, and Economics, ETH Zürich, Zürich, Switzerland

[2]Department of Emergency Medicine, Inselspital, Bern, University Hospital, University of Bern, Bern, Switzerland

[3]Department of Diabetes, Endocrinology, Nutritional Medicine and Metabolism, Inselspital, Bern, University Hospital, University of Bern, Bern, Switzerland

[4]Department of Endocrinology, Diabetes and Metabolic Diseases, Kantonsspital Olten, Olten, Switzerland

[5]Institute of AI in Management, LMU Munich, Munich, Germany

[6]Institute of Technology Management, University of St. Gallen, St. Gallen, Switzerland

[*]these authors contributed equally

Corresponding Author:
Aristomenis Exadaktylos, MD
Department of Emergency Medicine
Inselspital, Bern, University Hospital
University of Bern
Freiburgstrasse 16C
Bern, 3010
Switzerland
Phone: 41 31632244
Email: Aristomenis.Exadaktylos@insel.ch

Abstract

Background: To provide effective care for inpatients with COVID-19, clinical practitioners need systems that monitor patient health and subsequently allow for risk scoring. Existing approaches for risk scoring in patients with COVID-19 focus primarily on intensive care units (ICUs) with specialized medical measurement devices but not on hospital general wards.

Objective: In this paper, we aim to develop a risk score for inpatients with COVID-19 in general wards based on consumer-grade wearables (smartwatches).

Methods: Patients wore consumer-grade wearables to record physiological measurements, such as the heart rate (HR), heart rate variability (HRV), and respiration frequency (RF). Based on Bayesian survival analysis, we validated the association between these measurements and patient outcomes (ie, discharge or ICU admission). To build our risk score, we generated a low-dimensional representation of the physiological features. Subsequently, a pooled ordinal regression with time-dependent covariates inferred the probability of either hospital discharge or ICU admission. We evaluated the predictive performance of our developed system for risk scoring in a single-center, prospective study based on 40 inpatients with COVID-19 in a general ward of a tertiary referral center in Switzerland.

Results: First, Bayesian survival analysis showed that physiological measurements from consumer-grade wearables are significantly associated with patient outcomes (ie, discharge or ICU admission). Second, our risk score achieved a time-dependent area under the receiver operating characteristic curve (AUROC) of 0.73-0.90 based on leave-one-subject-out cross-validation.

Conclusions: Our results demonstrate the effectiveness of consumer-grade wearables for risk scoring in inpatients with COVID-19. Due to their low cost and ease of use, consumer-grade wearables could enable a scalable monitoring system.

Trial Registration: Clinicaltrials.gov NCT04357834; https://www.clinicaltrials.gov/ct2/show/NCT04357834

(JMIR Form Res 2022;6(6):e35717) doi:10.2196/35717

KEYWORDS

COVID-19; risk scoring; wearable devices; wearable; smartwatches; smartwatch; Bayesian survival analysis; remote monitoring; patient monitoring; remote patient monitoring; smart device; digital health; risk score; scalable; general ward; hospital; measurement tool; measurement instrument

Introduction

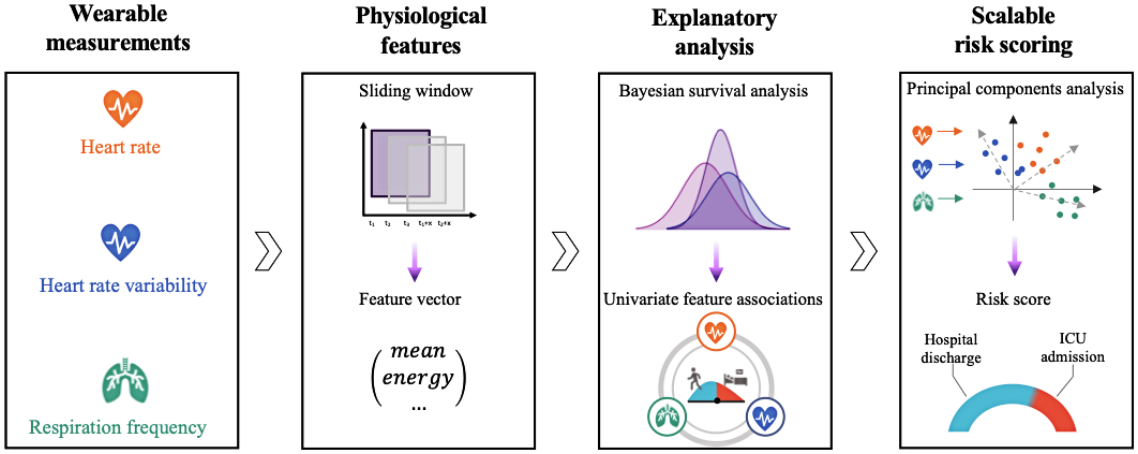
Health trajectories from patients with COVID-19 show large variability with sudden deterioration in the disease state and uncertain outcomes [1-4]. Hence, to provide effective care, clinical practitioners need systems that allow for monitoring the health trajectory of patients with COVID-19, especially during hospitalization [5-8]. Such systems can then be used to estimate the risk of a deterioration in the health condition and thus generate early warnings of critical conditions. In clinical practice, this enables the allocation of resources to patients in need and supports early responses to critical conditions [6,9-11].

Prior research has developed systems for monitoring patients with COVID-19 in different settings. One research stream detects the onset of COVID-19 using wearables (eg, smartphones) [12-14] and thus addresses the time before hospitalization. Another literature stream focuses on risk scoring for patients in intensive care units (ICUs) [7,8,15-17]. Here, monitoring systems are customized for the needs in intensive care and thus build upon specialized and often proprietary medical devices for physiological measurements. Vital signs, such as the heart rate (HR) or respiration frequency (RF), have been found to be predictive of critical health conditions [7,16].

In contrast, research is needed that develops systems for risk scoring for inpatients in general wards, which presents the focus of this work. This requires a custom risk score tailored to the corresponding patient population and nonspecialized monitoring devices that are available in general wards. For inpatients with COVID-19 in general wards, we propose the use of consumer-grade wearables (smartwatches) for monitoring and subsequent risk scoring due to their low cost, ease of use, and, thus, potential scalability. Previously, research has demonstrated the clinical relevance of consumer-grade wearables for longitudinal physiological measurements [18,19]. Further, they have been used for monitoring the progression of various other diseases (eg, diabetes mellitus [20,21]), yet their effectiveness for inpatients with COVID-19 in general wards remains to be confirmed.

In this paper, we develop a risk score for inpatients with COVID-19 in general wards based on scalable consumer-grade wearables (see Figure 1 for our overview). The consumer-grade wearables are used to monitor physiological measurements of patients: HR, heart rate variability (HRV), and RF. Based on these measurements, our risk score assesses the risk of different patient outcomes, defined as the probability of hospital discharge and ICU admission.

Figure 1. Monitoring and risk scoring. To develop a scalable risk score, physiological features were computed from wearable measurements. Next, Bayesian survival analysis was conducted to assess the association between the physiological features and patient outcomes. Lastly, a scalable risk score was developed. This study was designed to demonstrate the effectiveness of consumer-grade wearables for a scalable risk-scoring system in inpatients with COVID-19 in the general ward. ICU: intensive care unit.



Methods

Study Procedure

In visit 1 (V1), a study investigator explained the nature, purpose, and risks of the study and provided eligible patients with a copy of the patient information sheet. If written informed consent was obtained and eligibility criteria were met, the remaining screening information was obtained. A patient number was assigned to each patient in ascending order.

Eligible patients were provided with a Garmin vívoactive 4S (Garmin International Inc., Olathe, Kansas, USA) smartwatch and a Xiaomi Redmi 9 (Xiaomi Corp., Beijing, China) smartphone. Patients wore the wearable on the wrist of the dominant hand, if possible (otherwise the other hand).

After mounting the devices, the study investigator controlled the function of the devices and checked whether data transfer was working properly. In addition, the patient was instructed to fully charge both devices once per day or as needed. The smartwatch was worn during the whole study duration, that is,

from hospitalization in the general ward until the patient was admitted to the ICU or discharged home.

The study investigators were equipped with a monitoring dashboard allowing for observation of the charging status as well as functionality of the devices in use. If a patient was not capable of charging the devices themselves or the devices were not working properly, a member of the study team directly approached the patient and either charged the devices or solved possible technical issues.

In visit 2 (V2, close-out visit), the treating physician in the general ward informed the study team that 1 of the close-out criteria (admitted to the ICU or discharged home) had been met. A member of the study team then visited the patient and initiated the close-out visit. During V2, patients returned the wearable, the smartphone, and the charging cable. Completeness of data transfer to the back-end server was checked, and thereafter, all data on the devices were deleted.

Ethical Considerations

The study followed the Declaration of Helsinki, the guidelines of good clinical practice, Swiss health laws, and the ordinance on clinical research. The study was approved by the local ethics committee of Bern, Switzerland (ID 2020-00874). Each patient provided informed written consent before any study-related procedure.

Data Collection

The technical backbone of our data collection comprised 2 components: (1) a smartwatch that continuously collected physiological parameters and (2) a custom smartphone app to transfer the data to our server. In particular, the collected data were first transferred via Bluetooth to our self-developed smartphone app. Subsequently, the data were sent to a central database.

The smartwatch was used for measuring various physiological parameters. The recorded sensor measurements were the accelerometer (ACC), interbeat interval (IBI), HR, and RF. The ACC was sampled with 25 Hz. The HR and RF were logged once per minute. The IBI was recorded by logging the time of each heartbeat. The HR, IBI, and RF were derived from the photoplethysmography (PPG) sensor of the wearable.

Additional patient demographics (ie, patient age and sex) were collected by the clinical practitioners.

Data Processing

Data gathered from sensors embedded in consumer-grade wearable devices come along with inherent challenges for clinical usage. In particular, consumer smartwatches are by no means certified medical devices, and their sensor data may be subject to noise and missing values. We thus performed customized preprocessing of the sensor data as follows.

The HRV was computed based on a time series of IBIs. Of note, variability measures retrieved from an optical PPG signal should be referred to as pulse rate variability (PRV), whereas the variability measures retrieved from an electrocardiogram (ECG) should be referred to as the HRV. Since variability measures are significantly correlated, we followed the convention and speak here of the HRV [22-24]. First, measurement artifacts were filtered by removing IBIs that differed by more than 20% from the preceding IBI [25]. Furthermore, we used an adaptive threshold analysis for the HRV that discarded time windows with less than half of the expected heartbeats recorded by the measurement device [26]. This adaptive threshold prevents HRV values from being distorted due to insufficient data in a time window. Subsequently, the HRV in both the time domain and the frequency domain was calculated according to international guidelines [25]. For the frequency-domain features, one needs to estimate the power spectral density [27]. The time between 2 heartbeats changes. Hence, the IBI series is irregularly sampled. To avoid resampling, which bears the risk of distorted HRV features in case the proportion of missing data increases, we relied on the Lomb-Scargle method [28-32].

Feature Engineering

Our data showed substantial variation in the HR, HRV, and RF throughout the day, which was most likely due to changes in patient activity patterns. A confirmatory check showed strong dependence on the intensity of body movements throughout the day (see Multimedia Appendix 1). To compute daily physiological features that are robust against the activity patterns of patients and their biological rhythms, measurements taken during a time window from midnight to 5:00 a.m. each day were used. This time frame roughly corresponds to the phase of patients' night rest, as characterized by stable physiological measurements and minimal body movements (see Multimedia Appendix 1).

The wearable-based measurements of the HR, HRV, and RF were aggregated into a single value per time window using feature engineering. For the HR and RF, we computed 15 statistical features that reflect different properties of the distribution over time (eg, mean, skewness, SD). For the HRV, there exists an extensive amount of research on the effect of the window size on HRV features [33-36]. Here, we followed recommendations by Malik et al [25] and computed 19 time-domain and frequency-domain HRV features over intervals of 300 seconds before taking the mean over the full time window. The detailed list of features is provided in Multimedia Appendix 1. To ensure representativeness of the features, we required a sufficient number of valid measurements during the night—reasonably, minimum data of half of the measuring period during the night. After the application of all quality criteria (ie, IBI quality criteria and minimum coverage of the measurement period), 114 (69.1%) of 165 observations were retained. Here, 1 observation represents the aggregated physiological measurements of a patient from 1 specific night. Throughout the paper, the combination of wearable-based measurements (HR, HRV, RF) and feature engineering is referred to as *physiological features*.

For both preprocessing and feature engineering, we leveraged the publicly available Python package FLIRT, which is tailored to process wearable data [37]. By choosing the parameters as stated before, the entire pipeline can be reproduced.

Explanatory Analysis of the Association of Physiological Features With Patient Outcomes

To assess the association of physiological features with observed patient outcomes (ie, hospital discharge vs ICU admission), survival analysis was conducted. This allowed us to appropriately account for the time-to-event nature of the data and the presence of right censoring (see the Results section). Since the physiological features were updated each day, they were represented by time-dependent covariates in our survival analysis. Accordingly, a pooled regression approach [38,39] was chosen to flexibly account for the time dependence of the covariates. Moreover, as we did not observe an ICU readmission after a hospital discharge, both events should be regarded as competing risks and modeled via cause-specific hazards [40]. To make optimal use of the data in our study, both hazards for hospital discharge and ICU admission were estimated in a joint model. Specifically, the probabilities of hospital discharge, ICU admission, or no event (ie, continued stay) of patient "i" on day "t" were related to a regression function of the physiological features from the previous night. The probabilities were modeled jointly via an ordinal regression using a cumulative probability model with complementary log-log link [41-43]. This can also be interpreted as modeling the health condition of patients through a latent variable, where hospital discharge indicates a better health condition than continued stay and continued stay indicates a better health condition than ICU admission. An ordinal regression model accurately reflects this relationship, while offering high flexibility. Additionally, patient age and sex were considered demographic features. The model was specified in a fully Bayesian framework. Thereby, we ensured the robustness of our analysis by appropriately quantifying the uncertainty in parameter estimates. This is particularly important for limited sample sizes, as may likely be the case with newly emerged diseases. The formal specification of our model is provided in Multimedia Appendix 2. Of note, our approach has a particular connection with the well-known proportional hazards model [44] and can be interpreted as a Cox regression with time-dependent covariates that further accounts for competing risks in a joint model of hospital discharge and ICU admission probability.

In our explanatory analysis, we estimated univariate associations, which allowed us to identify the association of individual physiological features with patient health. Thus, a separate model was fitted for each physiological feature.

Development of a Risk Score

To develop a risk score based on the physiological features, a parsimonious 2-step approach was chosen. That is, we first used feature engineering and principal component analysis (PCA) to obtain a low-dimensional but comprehensive representation of patients' physiological state. This representation was then linked to patient outcomes through a Bayesian survival model that was similar to the models used in our explanatory analysis. We chose this approach over alternative methods (eg, deep learning) due to several reasons. First, the use of a parametric model can effectively reduce the risk of overfitting, while the feature engineering still allows us to use high-dimensional sensor data. Moreover, by jointly modeling the probabilities of hospital discharge and ICU admission, the risk score makes optimal use of the available data and can be readily interpreted as an overall indicator of patient condition. Finally, the use of Bayesian modeling ensures robust results even with limited amounts of data and appropriately quantifies uncertainty in the risk score.

The risk score was constructed by combining multiple physiological features into an overall metric. For this, we proceeded as follows: (1) The coefficients of the explanatory models were used to select physiological features of the HR, HRV, and RF that showed a relevant association (80% credible interval [CrI], excluding 0) with patient outcomes. (2) Since many features of the same measurement were strongly correlated, dimensionality reduction via PCA [45] was applied to generate a lower-dimensional representation of the underlying physiological information. (3) Pooled logistic least absolute shrinkage and selection operator (LASSO) regressions were employed to identify principal components (PCs) with high predictive power. Here, the PCs were used as predictors for the probability of either hospital discharge or ICU admission on a given day. The tuning parameter λ for the LASSO regularization was chosen via cross-validation. All PCs with nonzero coefficients were selected. (4) The risk score was computed from the linear predictor of a similar ordinal regression model as for the explanatory analysis but with the selected PCs as covariates. Correspondingly, a larger risk score implies a higher probability of ICU admission and a lower probability of hospital discharge. The probability of continued stay (ie, neither hospital discharge nor ICU admission) is $P_{\text{continued stay}} = 1 - P_{\text{discharge}} - P_{\text{ICU}}$.

Estimation and Performance Evaluation

All model parameters were estimated using a fully Bayesian framework [46-48]. Weakly informative priors were used for all parameters [49], and the estimation was checked by following best-practice recommendations in Bayesian modeling [46,50]. Details of the estimation and model checking are provided in Multimedia Appendices 2 and 3.

The performance of the developed risk score was evaluated via leave-one-patient-out cross-validation. The cross-validation covered all relevant preprocessing steps, including PCA. We assessed the performance in terms of discrimination accuracy via time-dependent receiver operating characteristic (ROC) curves using the incident case approach with dynamic controls (I/D) [51,52]. Moreover, in survival models with competing risks, ROC curves must be cause specific. Due to the small number of patients with ICU admission in our sample, we here focused on the ROC curve for hospital discharge. The ROC curve of the risk score–based prediction of hospital discharge for varying length of stay was thus evaluated to obtain a time-dependent area under the receiver operating characteristic curve (AUROC; see Multimedia Appendix 2 for details). The time-dependent AUROC assesses the predictive accuracy of the risk score to discriminate between patients who are discharged after a given number of days and patients who continue to stay in the hospital [52]. It can be interpreted as the probability that a random patient who is discharged on day "t"

has a higher predicted hazard of discharge than a random patient who continues to stay in the hospital [53].

To assess the added value of continuous physiological measurements for monitoring a patient's health throughout their hospital stay, we compared our risk score model to an alternative model that uses only data from the first night of hospital stay but is otherwise identical. The time-dependent AUROC was computed for both risk score models and compared across a varying length of stay in the hospital. We evaluated the length of stay for which a sufficient number of observations was available (ie, up to 6 days), corresponding to 87% of all observed lengths of stay. In the Results section, we further report a smoothed AUROC over time using the nearest-neighbor estimator for time-dependent ROC curves [54]. Our code is available in the official code repository [55].

Results

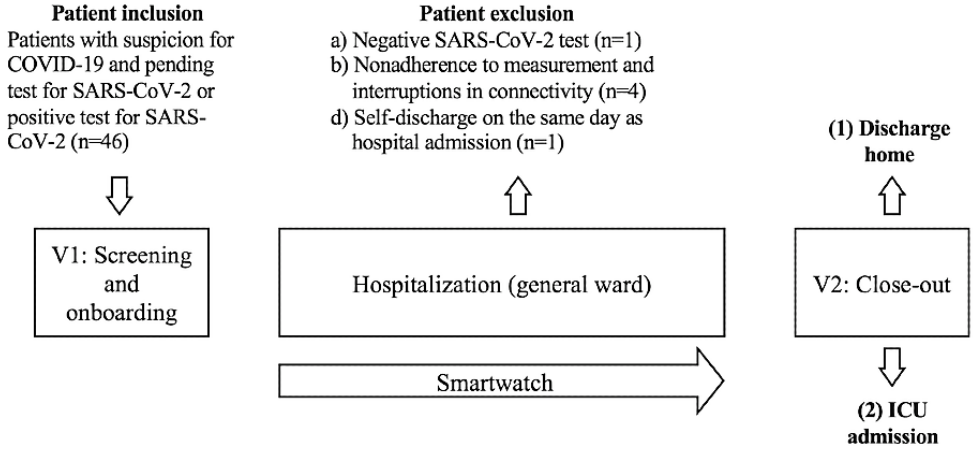
Study Setting

We conducted an observational study (see the study flowchart in Figure 2) between October 2020 and June 2021 in the general ward of a tertiary referral center in Switzerland. In total, 46 patients were recruited according to 2 different scenarios of recruitment and enrollment in the study: (1) Patients who attended the emergency ward and were hospitalized with suspicion of COVID-19 were recruited directly during their initial evaluation, and (2) additionally, all inpatients tested positive for SARS-CoV-2 were reported to the study team automatically with an email alert from the laboratory and thereafter contacted (in-hospital visit) by a member of the study team. Either of the following patient outcomes were possible: (1) hospital discharge or (2) ICU admission.

Inclusion criteria were age greater than 18 years, suspicion of COVID-19 or patient testing positive for SARS-CoV-2, and hospitalization in the general ward. Exclusion criteria were direct transfer from an emergency ward or external institution to the ICU (ie, no hospitalization in the general ward of the study institution). Further exclusion criteria were that the smartwatch could not be attached around the wrist of the patient, known allergies to components of the smartwatch, and rejection of ICU admission in the patient decree.

After screening, 1 (2.2%) of 46 individuals was excluded due to a negative SARS-CoV-2 result, 4 (8.7%) patients were excluded due to technical problems during the recording (eg, persistent interruptions of the Bluetooth connection between wearable and smartphone) or nonadherence to the prescribed measurement regime, and 1 (2.2%) individual was excluded because the hospital discharge occurred on the same day of hospitalization. In total, 40 (87%) patients remained. Of these, 7 (17.5%) were admitted to an ICU during their hospital stay (after a median of 2 days), and 31 (77.5%) were discharged without a subsequent ICU stay (after a median of 4 days). In addition, 2 (5%) patients dropped out before their outcome was recorded and were thus treated as right-censored in our analysis.

Figure 2. Overview of study with a study flowchart. Data were obtained according to the study flowchart. During visit 1 (V1), 46 eligible patients were recruited. After hospitalization in the general ward, patients were equipped with a consumer-grade wearable (smartwatch). We excluded patients with suspected COVID-19 in the case of a negative SARS-CoV-2 test (n=1). In addition, patients were excluded due to nonadherence to measurement principles or interruptions in connectivity (n=4) and self-discharge on the same day as hospital admission (n=1). During visit 2 (V2), we recorded the patient outcomes (ie, discharge, n=31, vs ICU admission, n=7). Patients with unknown outcomes were right-censored (n=2). ICU: intensive care unit.



Association of Physiological Features With Patient Outcomes

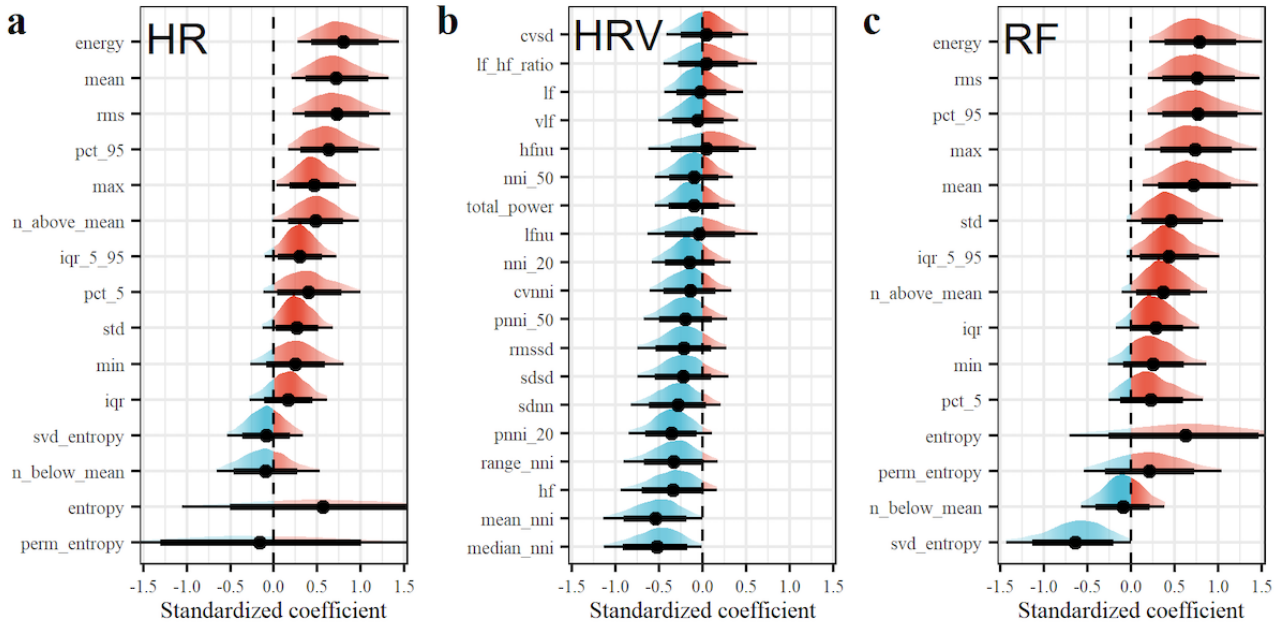
Figure 3 shows the association of physiological features with patient outcomes. Specifically, we reported standardized coefficients of the physiological features obtained from survival models adjusting for patient age and sex. A positive coefficient indicates that an increase in the value of a physiological feature on day "t" is associated with a higher probability of ICU admission as well as a lower probability of hospital discharge on day "t." In contrast, a negative coefficient indicates that an increase in the value of a physiological feature is associated with a lower probability of ICU admission and a higher probability of hospital discharge on a given day.

Overall, 49 different univariate associations were estimated (ie, with 15, 30.6%, features related to the HR, 19, 38.8%, features related to the HRV, and 15, 30.6%, features related to the RF). For features related to the HR, we found the following associations: a higher HR was associated with worsened patient outcomes. In particular, we found that an increase in the mean HR indicated a deterioration in the patient's condition (coefficient 0.71, 95% CrI 0.20-1.32). A similar association

was found for several other features, including the maximum HR (coefficient 0.46, 95% CrI 0.03-0.94). For entropy-based features, the estimated relationship remained largely uncertain, however. For features related to the HRV, we found that increases were associated with improved patient outcomes. For example, an increase in the standard deviation of normal-to-normal intervals (SDNN) indicated an improvement of the patient condition (coefficient –0.28, 95% CrI –0.82 to 0.21). Moreover, several features related to the RF showed a positive association, where larger values indicated a worsened patient outcome. For example, increases in the 95% quantile of the RF were associated with a deterioration in the patient's condition (coefficient 0.77, 95% CrI 0.19-1.51). The same was observed for the RF SD (0.46, 95% CrI –0.05 to 1.06). Altogether, these associations establish that the risk of a worsened condition among inpatients with COVID-19 can be identified through health measurements from consumer-grade wearables.

As part of our robustness checks, various alternative model specifications were tested (ie, changes with respect to the time window used for physiological measurements, the time trend, subject-specific variation, the cumulative distribution function, and wider priors; see Multimedia Appendix 4). We obtained similar estimates for all models, thus implying that the estimated associations between physiological features and patient outcomes remain robust.

Figure 3. Association of physiological features with patient outcomes. Shown are the standardized coefficients of physiological features for the (a) HR, (b) HRV, and (c) RF. Features were computed based on daily physiological measurements from wearables (see the Feature Engineering section and Multimedia Appendix 1). For each coefficient, we reported the posterior probability mass with mean (dot) and the 80% and 95% CrIs (thick and thin bars, respectively). Positive values (red) indicate an association with a deterioration in the health condition, and negative values (blue) indicate an association with an improved health condition. CrI: credible interval; HR: heart rate; HRV: heart rate variability; RF: respiration frequency.

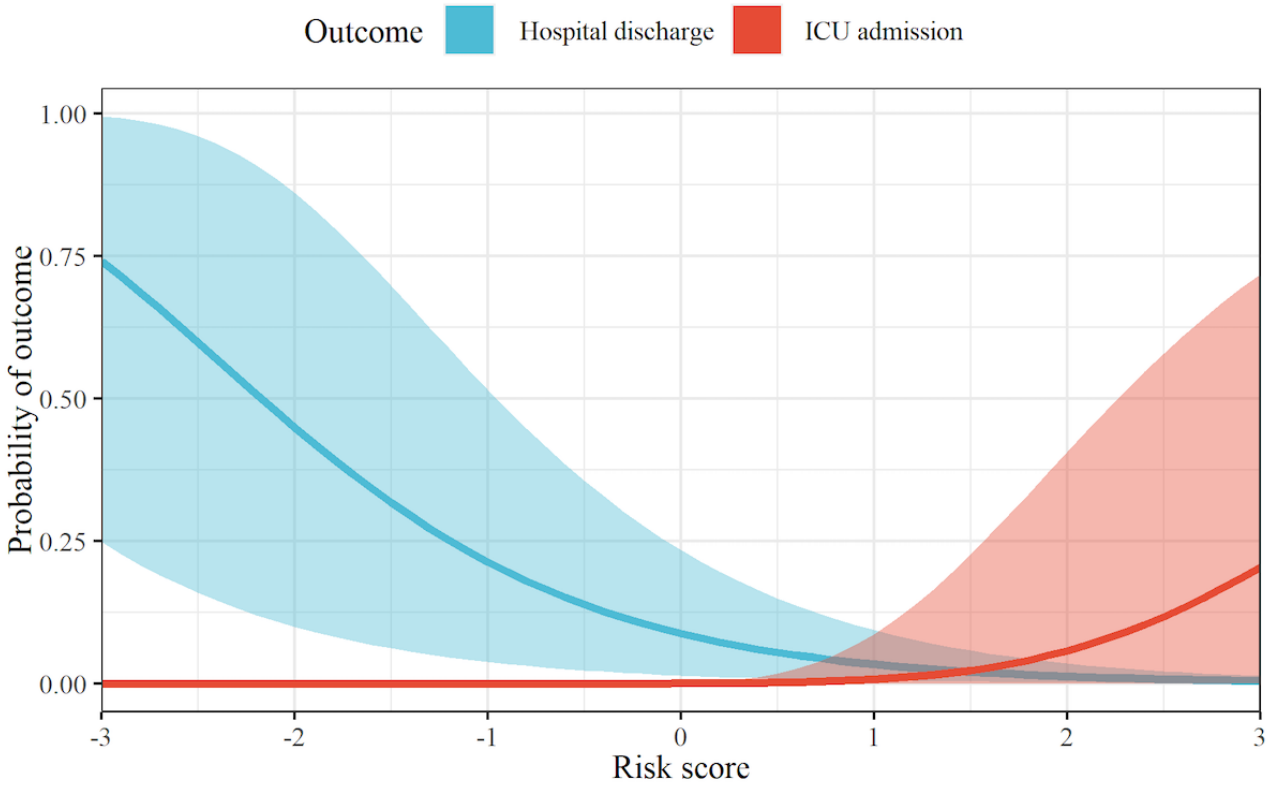


Development of a Risk Score

Next, the physiological measurements from the wearables were combined into an overall risk score. Here, our main aim was to demonstrate that a combination of different physiological features is of predictive value and thus jointly informative. For this, PCA [45] was applied to all features that showed a relevant association (80% CrI, excluding 0) with patient outcomes. This was the case for 9 HR features, 4 HRV features, and 9 RF features. Next, PCs with the highest predictive value for patient outcomes were identified using LASSO (see the Methods section). For our clinical data, the LASSO selected 8 (36.4%) of 22 PCs. These PCs characterized the physiological state of patients through a lower-dimensional representation of the wearable-based measurements. A visualization of the PCs is shown in Multimedia Appendix 5.

The selected PCs were used to model patient outcomes as the dependent variable based on a survival model that is similar to that of the explanatory analysis. Multimedia Appendix 5 reports the estimated coefficients. The resulting linear predictor for the probability of hospital discharge and ICU admission was used as the overall risk score. The risk score thus quantifies the probability of hospital discharge and ICU admission of patients on a given day using wearable-based measurements from the previous night. Here, a higher score generally indicates a worse patient condition (Figure 4). Although the overall health condition of patients is confidently predicted by the risk score, the smaller number of patients with ICU admission in our data set means that the risk score is less differentiated with regard to ICU admission.

Figure 4. Probability of hospital discharge and ICU admission for different values of the risk score. Shown is the estimated daily probability of hospital discharge (blue) and ICU admission (red) as a function of the risk score. A larger risk score implies a higher probability of ICU admission and a lower probability of hospital discharge. Posterior means (lines) and 95% CrIs (shaded areas) are reported. The probability of continued stay (ie, neither hospital discharge nor ICU admission) is not shown but can be computed as $P_{\text{continued stay}} = 1 - P_{\text{discharge}} - P_{\text{ICU}}$. CrI: credible interval; ICU: intensive care unit.



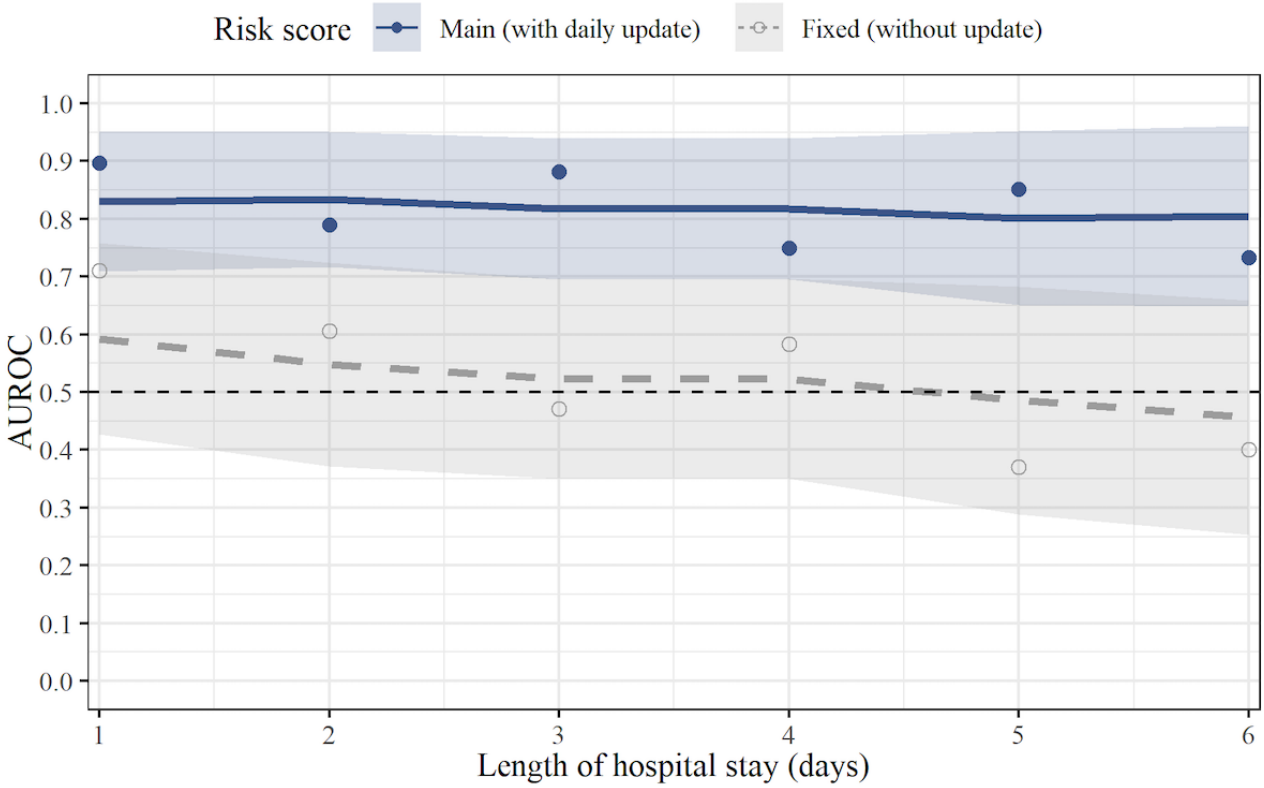
Evaluation of the Risk Score

Figure 5 shows the leave-one-patient-out cross-validation results for the predictive performance of our risk score. Because the risk score was updated as the condition of patients changed throughout their hospital stay, the time-dependent AUROC was used as a performance metric that accounts for time-varying prediction performance. A daily AUROC was computed for up to 6 days of hospital stay, which covered 87% of the patients' length of stay. For different lengths of stay, the risk score achieved a time-dependent AUROC of 0.73-0.90, suggesting reasonable predictive performance (Figure 5). This establishes that the different physiological features are jointly informative of patient health condition over time.

For comparison, we also reported the performance of a fixed risk score that used only data from the first night of hospital stay but was otherwise identical (Figure 5). Comparing the performance of the fixed risk score and our original risk score allowed us to assess the benefit of daily updating the physiological measurements. For the first day of hospital stay, the fixed risk score achieved a performance that was worse than the risk score with updated physiological measurements but was still above 0.70. However, for a length of stay longer than 1 day, the fixed risk score showed a consistently inferior performance.

To further assess the added value of the physiological features used in our risk score, we also evaluated the performance of a risk score that uses demographic features (ie, patient age and sex) but no physiological features. The cross-validation results for this risk score indicated that demographic features alone have no relevant predictive value with regard to the time-varying health condition of the patients in our sample (see Multimedia Appendix 6). Together, these results confirm that continuously updated, repeated monitoring of physiological measurements can provide an added value for analyzing the patient's condition during the hospital stay.

Figure 5. Prediction performance of the risk score over time. Shown is the time-dependent AUROC of the risk score in predicting patient discharge over time. Two scenarios are compared: (1) main (blue solid line) and (2) fixed (gray dashed line). In the main scenario, the daily risk score is computed from updated wearable-based measurements recorded during the respective previous night. The AUROC is significantly above 0.5 for up to 6 days, which covers 87% of the patients' length of stay. In the fixed scenario, the risk score is computed throughout the stay from recordings only from the first night. The comparison between these scenarios shows the added value of regularly updated health measurements provided by wearables. Out-of-sample predictions were obtained via leave-one-patient-out cross-validation. Dots show the individual time-dependent AUROC estimates for days with observed patient discharge. Smoothing was performed via a nearest-neighbor estimator (see the Performance Evaluation section) to obtain an estimate of the mean AUROC over time (lines) with 95% CIs (shaded areas). AUROC: area under the receiver operating characteristic curve.



Discussion

Principal Results

This work presents a monitoring system that allows for risk scoring of inpatients with COVID-19 in the general ward using consumer-grade wearables (smartwatches).

For this, Bayesian survival analysis was used to establish that physiological measurements monitored by consumer-grade wearables are indicative of patient outcomes in the general ward (ie, hospital discharge vs ICU admission). We further showed that these different physiological measurements can be combined into a single, clinically meaningful risk score with high prediction performance regarding the health condition of patients (time-dependent AUROC of 0.73-0.90). Our results show the feasibility of a risk score for inpatients with COVID-19 in general wards based on scalable consumer-grade wearables. In the future, such risk scores may enable clinical practitioners to adapt to patient needs and, ideally, respond earlier when a patient trajectory progresses toward a critical condition.

We found that several physiological features derived from wearable-based measurements are associated with patient outcomes. For instance, a higher mean HR, a higher mean RF, and a lower HRV RMSSD are all indicative of a deterioration in the health condition of patients. The observed relationship between patient outcomes and cardiovascular features (HR and HRV), as well as patient outcomes and RF measurements, is consistent with previous research on digital biomarkers [22,56-60]. These findings add to the robustness of our monitoring system. Importantly, we discovered these associations based on consumer-grade wearables, which indicates the clinical applicability and, thus, the relevance of the technology. Furthermore, the risk score may implicitly capture information on clinical interventions (eg, ventilation, which affects the RF). Hence, wearable recordings must be interpreted carefully in the light of other simultaneous interventions.

To derive our risk score, we intentionally chose a parsimonious approach using feature engineering and Bayesian survival modeling. Different from other machine learning methods, a parametric, Bayesian approach like ours is especially viable in the case of newly emerged diseases, where data availability may be limited. Since our feature engineering is mostly disease independent, the physiological features could be integrated into models for other diseases too, further promoting scalability. More generally, our approach demonstrates how multiple competing patient outcomes can be flexibly linked to time-dependent measurements in a parametric, joint model of patient condition.

Comparison With Prior Work

Prior research successfully explored vital signs measured by smartwatches (eg, the resting HR) as a basis to detect the onset of COVID-19 outside a clinical setting [12-14]. Hence, we leveraged similar devices to record physiological measures and model a similar outcome (ie, deterioration in a patient's health). However, our monitoring system differs from others on COVID-19 as follows: First, there is no proof-of-concept study in a clinical setup that explores smartwatches as a basis to monitor patients with COVID-19 to the best of our knowledge. Second, we modeled a patient's health condition as a whole to detect not only a deterioration in the patient's health but also an improvement.

Lastly, several studies have focused on risk scoring in ICUs [7,8,16,17]. However, due to a large number of hospitalizations for COVID-19, inpatients in general wards are also of major concern. Different from our study setting, risk scoring in ICUs builds upon specialized medical devices for health monitoring and a specific patient population. Because of this, a direct transfer of ICU risk scores to clinical practice in general wards is obviously limited. Therefore, we developed a monitoring system and subsequent risk scoring that is particularly suited for general wards (eg, there is no need for specialized medical monitoring technology).

In summary, our study supports the clinical relevance of wearables exclusively based on consumer-grade technology. In contrast to specialized medical devices for health monitoring (eg, finger pulse oximetry or ECG sensor), consumer-grade technology comes at a comparatively low cost, can be deployed easily, and is thus scalable. Clinical practitioners simply need to attach the smartwatch to the wrist of a patient. In addition, smartwatches offer a familiar user interface.

Limitations

A general concern may be that measurements from consumer-grade wearables are subject to noise or missing values. The results of our study, however, show that a wearable-based risk score can offer robust predictions of patient outcomes. Our study opens several possibilities for future research. The main limiting factor of our study is the sample size of 40, which naturally restricts the number of ICU admissions in the data set. To further assess the predictive performance of wearable-based risk scores, in particular with regard to ICU admission, future research might expand our data set with additional patient populations and different variants of SARS-CoV-2. The model merely incorporated data from wearable sensors for risk scoring and refrained from integrating other data sources (eg, electronic health records). This choice was made to ensure a scalable use in clinical practice. Further, our system builds upon dimensionality reduction via PCA to handle high-dimensional sensor data, proving effective to avoid overfitting. Nevertheless, future research may explore alternative machine learning methods for risk scoring.

Conclusion

Overall, our results show the promise of consumer-grade wearables as an effective, scalable, and low-cost technology for health monitoring in a general ward. In the future, consumer-grade wearables, such as smartwatches, may further offer monitoring capabilities for inpatients with other diseases.

Acknowledgments

S Föll, MM, KK, VL, TZ, DS, SJ, and FW contributed to the conception and design of the study. MM and S Föll developed the digital biomarker platform. KK, SJ, and DS screened and enrolled the patients and acquired the data. MM and S Föll processed the data. AL conducted the statistical analysis. S Feuerriegel supervised the statistical analysis. AE supervised the clinical study. S Föll and AL wrote the manuscript. MM, FW, EF, S Feuerriegel, KK, VL, TZ, DS, SJ, and AE critically reviewed and edited the manuscript. FW and AE jointly supervised the project and share the last authorship. All authors have approved the final draft of the manuscript for submission.

This work was partially funded by the Hasler Stiftung (Project 20039). The funding body had no control over the study design, data collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication.

Conflicts of Interest

S Feuerriegel declares membership in a COVID-19 working group of the World Health Organization (WHO) but without competing interests. All other authors declare no competing interest.

Multimedia Appendix 1
Data description.
[DOCX File , 63 KB - formative_v6i6e35717_app1.docx]

Multimedia Appendix 2
Method details.
[DOCX File , 28 KB - formative_v6i6e35717_app2.docx]

Multimedia Appendix 3
Model Diagnostics.

[DOCX File , 111 KB - formative_v6i6e35717_app3.docx]

Multimedia Appendix 4
Robustness of Explanatory Analysis.
[DOCX File , 156 KB - formative_v6i6e35717_app4.docx]

Multimedia Appendix 5
Principal component analysis (PCA) results.
[DOCX File , 120 KB - formative_v6i6e35717_app5.docx]

Multimedia Appendix 6
Comparison to a Risk Score Using Only Demographic Features.
[DOCX File , 54 KB - formative_v6i6e35717_app6.docx]

References

1. Bolourani S, Brenner M, Wang P, McGinn T, Hirsch JS, Barnaby D, Northwell COVID-19 Research Consortium. A machine learning prediction model of respiratory failure within 48 hours of patient admission for COVID-19: model development and validation. J Med Internet Res 2021 Feb 10;23(2):e24246. [doi: 10.2196/24246] [Medline: 33476281]
2. García LF. Immune response, inflammation, and the clinical spectrum of COVID-19. Front Immunol 2020 Jun 16;11:1441 [FREE Full text] [doi: 10.3389/fimmu.2020.01441] [Medline: 32612615]
3. Jose RJ, Manuel A. COVID-19 cytokine storm: the interplay between inflammation and coagulation. Lancet Respir Med 2020 Jun;8(6):e46-e47. [doi: 10.1016/s2213-2600(20)30216-2]
4. Razavian N, Major VJ, Sudarshan M, Burk-Rafel J, Stella P, Randhawa H, et al. A validated, real-time prediction model for favorable outcomes in hospitalized COVID-19 patients. NPJ Digit Med 2020 Oct 06;3(1):130 [FREE Full text] [doi: 10.1038/s41746-020-00343-x] [Medline: 33083565]
5. Cummings BC, Ansari S, Motyka JR, Wang G, Medlin RP, Kronick SL, et al. Predicting intensive care transfers and other unforeseen events: analytic model validation study and comparison to existing methods. JMIR Med Inform 2021 Apr 21;9(4):e25066 [FREE Full text] [doi: 10.2196/25066] [Medline: 33818393]
6. Schwab P, Mehrjou A, Parbhoo S, Celi LA, Hetzel J, Hofer M, et al. Real-time prediction of COVID-19 related mortality using electronic health records. Nat Commun 2021 Feb 16;12(1):1058 [FREE Full text] [doi: 10.1038/s41467-020-20816-7] [Medline: 33594046]
7. Zhao Z, Chen A, Hou W, Graham JM, Li H, Richman PS, et al. Prediction model and risk scores of ICU admission and mortality in COVID-19. PLoS One 2020;15(7):e0236618 [FREE Full text] [doi: 10.1371/journal.pone.0236618] [Medline: 32730358]
8. Subudhi S, Verma A, Patel AB, Hardin CC, Khandekar MJ, Lee H, et al. Comparing machine learning algorithms for predicting ICU admission and mortality in COVID-19. NPJ Digit Med 2021 May 21;4(1):87 [FREE Full text] [doi: 10.1038/s41746-021-00456-x] [Medline: 34021235]
9. Cho S, Park S, Song M, Bae YY, Lee D, Kim D. Prognosis score system to predict survival for COVID-19 cases: a Korean nationwide cohort study. J Med Internet Res 2021 Feb 22;23(2):e26257 [FREE Full text] [doi: 10.2196/26257] [Medline: 33539312]
10. Emanuel EJ, Persad G, Upshur R, Thome B, Parker M, Glickman A, et al. Fair allocation of scarce medical resources in the time of Covid-19. N Engl J Med 2020 May 21;382(21):2049-2055 [FREE Full text] [doi: 10.1056/NEJMsb2005114] [Medline: 32202722]
11. Truog RD, Mitchell C, Daley GQ. The toughest triage - allocating ventilators in a pandemic. N Engl J Med 2020 May 21;382(21):1973-1975. [doi: 10.1056/NEJMp2005689] [Medline: 32202721]
12. Quer G, Radin JM, Gadaleta M, Baca-Motes K, Ariniello L, Ramos E, et al. Wearable sensor data and self-reported symptoms for COVID-19 detection. Nat Med 2021 Jan 29;27(1):73-77. [doi: 10.1038/s41591-020-1123-x] [Medline: 33122860]
13. Mishra T, Wang M, Metwally AA, Bogu GK, Brooks AW, Bahmani A, et al. Pre-symptomatic detection of COVID-19 from smartwatch data. Nat Biomed Eng 2020 Dec 18;4(12):1208-1220 [FREE Full text] [doi: 10.1038/s41551-020-00640-6] [Medline: 33208926]
14. Menni C, Valdes AM, Freidin MB, Sudre CH, Nguyen LH, Drew DA, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. Nat Med 2020 Jul 11;26(7):1037-1040 [FREE Full text] [doi: 10.1038/s41591-020-0916-2] [Medline: 32393804]
15. Pan P, Li Y, Xiao Y, Han B, Su L, Su M, et al. Prognostic assessment of COVID-19 in the intensive care unit by machine learning methods: model development and validation. J Med Internet Res 2020 Nov 11;22(11):e23128 [FREE Full text] [doi: 10.2196/23128] [Medline: 33035175]

16. Cheng F, Joshi H, Tandon P, Freeman R, Reich DL, Mazumdar M, et al. Using machine learning to predict ICU transfer in hospitalized COVID-19 patients. J Clin Med 2020 Jun 01;9(6):1668 [FREE Full text] [doi: 10.3390/jcm9061668] [Medline: 32492874]
17. Galloway JB, Norton S, Barker RD, Brookes A, Carey I, Clarke BD, et al. A clinical risk score to identify patients with COVID-19 at high risk of critical care admission or death: an observational cohort study. J Infect 2020 Aug;81(2):282-288 [FREE Full text] [doi: 10.1016/j.jinf.2020.05.064] [Medline: 32479771]
18. Nelson BW, Allen NB. Accuracy of consumer wearable heart rate measurement during an ecologically valid 24-hour period: intraindividual validation study. JMIR Mhealth Uhealth 2019 Mar 11;7(3):e10828 [FREE Full text] [doi: 10.2196/10828] [Medline: 30855232]
19. Dunn J, Kidzinski L, Runge R, Witt D, Hicks JL, Schüssler-Fiorenza Rose SM, et al. Wearable sensors enable personalized predictions of clinical laboratory measurements. Nat Med 2021 Jun 24;27(6):1105-1112 [FREE Full text] [doi: 10.1038/s41591-021-01339-0] [Medline: 34031607]
20. Bent B, Cho PJ, Henriquez M, Wittmann A, Thacker C, Feinglos M, et al. Engineering digital biomarkers of interstitial glucose from noninvasive smartwatches. NPJ Digit Med 2021 Jun 02;4(1):89 [FREE Full text] [doi: 10.1038/s41746-021-00465-w] [Medline: 34079049]
21. Maritsch M, Föll S, Lehmann V, Bérubé C, Kraus M, Feuerriegel S, et al. Towards wearable-based hypoglycemia detection and warning in diabetes. 2020 Apr Presented at: CHI EA '20: Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems; April 2020; Honolulu, HI p. 1-8 URL: https://linkinghub.elsevier.com/retrieve/pii/S0169-2607(21)00535-6 [doi: 10.1145/3334480.3382808]
22. Natarajan A, Su H, Heneghan C. Assessment of physiological signs associated with COVID-19 measured using wearable devices. NPJ Digit Med 2020 Nov 30;3(1):156 [FREE Full text] [doi: 10.1038/s41746-020-00363-7] [Medline: 33299095]
23. Orphanidou C. Signal Quality Assessment in Physiological Monitoring. State of the Art and Practical Considerations. Berlin: Springer-Verlag; 2018.
24. Lu S, Zhao H, Ju K, Shin K, Lee M, Shelley K, et al. Can photoplethysmography variability serve as an alternative approach to obtain heart rate variability information? J Clin Monit Comput 2008 Feb;22(1):23-29 [FREE Full text] [doi: 10.1007/s10877-007-9103-y] [Medline: 17987395]
25. Malik M, Bigger JT, Camm AJ, Kleiger RE, Malliani A, Moss AJ, et al. Heart rate variability: standards of measurement, physiological interpretation, and clinical use. Eur Heart J 1996 Mar 01;17(3):354-381. [doi: 10.1093/oxfordjournals.eurheartj.a014868]
26. Bent B, Goldstein BA, Kibbe WA, Dunn JP. Investigating sources of inaccuracy in wearable optical heart rate sensors. NPJ Digit Med 2020;3:18 [FREE Full text] [doi: 10.1038/s41746-020-0226-6] [Medline: 32047863]
27. Bachler M. Spectral analysis of unevenly spaced data: models and application in heart rate variability. SNE 2017 Dec;27(4):183-190. [doi: 10.11128/sne.27.tn.10393]
28. Schaffer T, Hensel B, Weigand C, Schüttler J, Jeleazcov C. Evaluation of techniques for estimating the power spectral density of RR-intervals under paced respiration conditions. J Clin Monit Comput 2014 Oct 19;28(5):481-486. [doi: 10.1007/s10877-013-9447-4] [Medline: 23508826]
29. Laguna P, Moody G, Mark R. Power spectral density of unevenly sampled data by least-square analysis: performance and application to heart rate signals. IEEE Trans Biomed Eng 1998 Jun;45(6):698-715. [doi: 10.1109/10.678605] [Medline: 9609935]
30. Morelli D, Rossi A, Cairo M, Clifton DA. Analysis of the impact of interpolation methods of missing RR-intervals caused by motion artifacts on HRV features estimations. Sensors (Basel) 2019 Jul 18;19(14):3163 [FREE Full text] [doi: 10.3390/s19143163] [Medline: 31323850]
31. Lomb NR. Least-squares frequency analysis of unequally spaced data. Astrophys Space Sci 1976 Feb;39(2):447-462. [doi: 10.1007/bf00648343]
32. Moody G. Spectral analysis of heart rate without resampling. 1993 Presented at: 10th Annual Conference Proceedings of Computers in Cardiology Conference; 1993; Aachen, Germany p. 715-718. [doi: 10.1109/cic.1993.378302]
33. Acharya RU, Joseph PK, Kannathal N, Lim CM, Suri JS. Heart rate variability: a review. Med Biol Eng Comput 2006 Dec 17;44(12):1031-1051. [doi: 10.1007/s11517-006-0119-0] [Medline: 17111118]
34. Shaffer F, Ginsberg JP. An overview of heart rate variability metrics and norms. Front Public Health 2017 Sep;5:258 [FREE Full text] [doi: 10.3389/fpubh.2017.00258] [Medline: 29034226]
35. Spiers JP, Silke B, McDermott U, Shanks RG, Harron DWG. Time and frequency domain assessment of heart rate variability: a theoretical and clinical appreciation. Clin Auton Res 1993 Apr;3(2):145-158. [doi: 10.1007/bf01819000]
36. Mietus JE, Peng CK, Henry I, Goldsmith RL, Goldberger AL. The pNNx files: re-examining a widely used heart rate variability measure. Heart 2002 Oct 11;88(4):378-380 [FREE Full text] [doi: 10.1136/heart.88.4.378] [Medline: 12231596]
37. Föll S, Maritsch M, Spinola F, Mishra V, Barata F, Kowatsch T, et al. FLIRT: a feature generation toolkit for wearable data. Comput Methods Programs Biomed 2021 Nov;212:106461 [FREE Full text] [doi: 10.1016/j.cmpb.2021.106461] [Medline: 34736174]

38. D'Agostino RB, Lee M, Belanger AJ, Cupples LA, Anderson K, Kannel WB. Relation of pooled logistic regression to time dependent Cox regression analysis: the Framingham Heart Study. Stat Med 1990 Dec;9(12):1501-1515. [doi: 10.1002/sim.4780091214] [Medline: 2281238]
39. Ngwa JS, Cabral HJ, Cheng DM, Pencina MJ, Gagnon DR, LaValley MP, et al. A comparison of time dependent Cox regression, pooled logistic regression and cross sectional pooling with simulations and an application to the Framingham Heart Study. BMC Med Res Methodol 2016 Nov 03;16(1):148 [FREE Full text] [doi: 10.1186/s12874-016-0248-6] [Medline: 27809784]
40. Austin PC, Lee DS, Fine JP. Introduction to the analysis of survival data in the presence of competing risks. Circulation 2016 Feb 09;133(6):601-609. [doi: 10.1161/circulationaha.115.017719]
41. Agresti A. Analysis of Ordinal Categorical Data, Second Edition. Hoboken, NJ: John Wiley & Sons; 2010.
42. McCullagh P. Regression models for ordinal data. J R Stats Soc (Methodol) 2018 Dec 05;42(2):109-127. [doi: 10.1111/j.2517-6161.1980.tb01109.x]
43. Bürkner P, Vuorre M. Ordinal regression models in psychology: a tutorial. Adv Methods Pract Psychol Sci 2019 Feb 25;2(1):77-101. [doi: 10.1177/2515245918823199]
44. Cox DR. Regression models and life-tables. J R Stats Soc (Methodol) 2018 Dec 05;34(2):187-202. [doi: 10.1111/j.2517-6161.1972.tb00899.x]
45. Hotelling H. Analysis of a complex of statistical variables into principal components. J Educ Psychol 1933;24(6):417-441. [doi: 10.1037/h0071325]
46. Gelman A, Rubin D, Carlin J, Stern H. Bayesian Data Analysis. 1st ed. London, UK: Chapman and Hall/CRC; 2021.
47. Bürkner P. brms: an R package for Bayesian multilevel models using Stan. J Stats Softw 2017;80(1):1-28. [doi: 10.18637/jss.v080.i01]
48. Hoffman M, Gelman A. The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. J Mach Learn Res 2014;15(1):1593-1623.
49. Stan Development Team. Stan Language Reference Manual (Version 2.25). URL: https://mc-stan.org/docs/2_21/reference-manual [accessed 2022-06-01]
50. Stan Development Team. Prior Choice Recommendations. URL: https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations [accessed 2022-06-01]
51. Heagerty PJ, Zheng Y. Survival model predictive accuracy and ROC curves. Biometrics 2005 Mar;61(1):92-105. [doi: 10.1111/j.0006-341x.2005.030814.x]
52. Saha P, Heagerty PJ. Time-dependent predictive accuracy in the presence of competing risks. Biometrics 2010 Dec;66(4):999-1011 [FREE Full text] [doi: 10.1111/j.1541-0420.2009.01375.x] [Medline: 20070296]
53. Bansal A, Heagerty PJ. A tutorial on evaluating the time-varying discrimination accuracy of survival models used in dynamic decision making. Med Decis Making 2018 Oct 14;38(8):904-916. [doi: 10.1177/0272989x18801312]
54. Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. Biometrics 2000 Jun;56(2):337-344. [doi: 10.1111/j.0006-341x.2000.00337.x] [Medline: 10877287]
55. Official code repository. 2021. URL: https://github.com/im-ethz/wave [accessed 2022-06-01]
56. Bhatraju PK, Ghassemieh BJ, Nichols M, Kim R, Jerome KR, Nalla AK, et al. Covid-19 in critically ill patients in the Seattle region - case series. N Engl J Med 2020 May 21;382(21):2012-2022 [FREE Full text] [doi: 10.1056/NEJMoa2004500] [Medline: 32227758]
57. Hirten RP, Danieletto M, Tomalin L, Choi KH, Zweig M, Golden E, et al. Use of physiological data from a wearable device to identify SARS-CoV-2 infection and symptoms and predict COVID-19 diagnosis: observational study. J Med Internet Res 2021 Feb 22;23(2):e26107 [FREE Full text] [doi: 10.2196/26107] [Medline: 33529156]
58. Hasty F, García G, Dávila CH, Wittels S, Hendricks S, Chong S. Heart rate variability as a possible predictive marker for acute inflammatory response in COVID-19 patients. Mil Med 2020 Nov 18;186(1-2):e34-e38 [FREE Full text] [doi: 10.1093/milmed/usaa405] [Medline: 33206183]
59. Massaroni C, Nicolò A, Schena E, Sacchetti M. Remote respiratory monitoring in the time of COVID-19. Front Physiol 2020 May 29;11:635 [FREE Full text] [doi: 10.3389/fphys.2020.00635] [Medline: 32574240]
60. Xu J, Wang W, Ye H, Pang W, Pang P, Tang M, et al. A predictive score for progression of COVID-19 in hospitalized persons: a cohort study. NPJ Prim Care Respir Med 2021 Jun 03;31(1):33 [FREE Full text] [doi: 10.1038/s41533-021-00244-w] [Medline: 34083541]

Abbreviations

ACC: accelerometer
AUROC: area under the receiver operating characteristic curve
CrI: credible interval
ECG: electrocardiogram
HRV: heart rate variability
HR: heart rate

HRV: heart rate variability
IBI: interbeat interval
ICU: intensive care unit
LASSO: least absolute shrinkage and selection operator
PC: principal component
PCA: principal component analysis
PPG: photoplethysmography
RF: respiration frequency
ROC: receiver operating characteristic

Edited by A Mavragani; submitted 15.12.21; peer-reviewed by J Dunn, C Smeets; comments to author 14.02.22; revised version received 06.04.22; accepted 09.05.22; published 21.06.22.

Please cite as:
Föll S, Lison A, Maritsch M, Klingberg K, Lehmann V, Züger T, Srivastava D, Jegerlehner S, Feuerriegel S, Fleisch E, Exadaktylos A, Wortmann F
A Scalable Risk-Scoring System Based on Consumer-Grade Wearables for Inpatients With COVID-19: Statistical Analysis and Model Development
JMIR Form Res 2022;6(6):e35717
URL: https://formative.jmir.org/2022/6/e35717
doi:10.2196/35717
PMID:35613417

Original Paper

Digital Screening and Automated Resource Identification System to Address COVID-19–Related Behavioral Health Disparities: Feasibility Study

Colleen Stiles-Shields[1], PhD; Kathryn R Batts[2], PhD; Karen M Reyes[1], BA; Joseph Archer[1,3], MD, MPH; Sharad Crosby[1,4], MS; Janel M Draxler[1], MSN; Nia Lennan[1], BA; Philip Held[1], PhD

[1]Section of Community Behavioral Health, Department of Psychiatry and Behavioral Sciences, Rush University Medical Center, Chicago, IL, United States

[2]Community Health Research Division, RTI International, Research Triangle Park, NC, United States

[3]University of Wisconsin School of Medicine and Public Health, Madison, WI, United States

[4]Institute for Translational Medicine, University of Chicago, Chicago, IL, United States

Corresponding Author:
Colleen Stiles-Shields, PhD
Section of Community Behavioral Health
Department of Psychiatry and Behavioral Sciences
Rush University Medical Center
1645 W Jackson Blvd
Suite 302
Chicago, IL, 60612
United States
Phone: 1 312 942 9998
Email: colleen_stiles-shields@rush.edu

Abstract

Background: Digital mental health (DMH) tools use technology (eg, websites and mobile apps) to conveniently deliver mental health resources to users in real time, reducing access barriers. Underserved communities facing health care provider shortages and limited mental health resources may benefit from DMH tools, as these tools can help improve access to resources.

Objective: This study described the development and feasibility evaluation of the Emotional Needs Evaluation and Resource Guide for You (ENERGY) System, a DMH tool to meet the mental health and resource needs of youth and their families developed in the context of the COVID-19 pandemic. The ENERGY System offers a brief assessment of resource needs; problem-solving capabilities; and symptoms of depression, anxiety, trauma, and alcohol and substance use followed by automated, personalized feedback based on the participant's responses.

Methods: Individuals aged ≥15 years were recruited through community partners, community events, targeted electronic health record messages, and social media. Participants completed screening questions to establish eligibility, entered demographic information, and completed the ENERGY System assessment. Based on the participant's responses, the ENERGY System immediately delivered digital resources tailored to their identified areas of need (eg, relaxation). A subset of participants also voluntarily completed the following: COVID-19 Exposure and Family Impact Survey (CEFIS) or COVID-19 Exposure and Family Impact Survey Adolescent and Young Adult Version (CEFIS-AYA); resource needs assessment; and feedback on their experience using the ENERGY System. If resource needs (eg, housing and food insecurity) were endorsed, lists of local resources were provided.

Results: A total of 212 individuals accessed the ENERGY System link, of which 96 (45.3%) completed the screening tool and 86 (40.6%) received resources. Participant responses on the mental health screening questions triggered on average 2.04 (SD 1.94) intervention domains. Behavioral Activation/Increasing Activities was the most frequently launched intervention domain (56%, 54/96), and domains related to alcohol or substance use were the least frequent (4%, 4/96). The most frequently requested support areas were finances (33%, 32/96), transportation (26%, 25/96), and food (24%, 23/96). The CEFIS and CEFIS-AYA indicated higher than average impacts from the pandemic (ie, average scores >2.5). Participants were satisfied with the ENERGY System overall (65%, 39/60) as well as the length of time it took to answer the questions (90%, 54/60), which they found easy to answer (87%, 52/60).

Conclusions: This study provided initial support for the feasibility of the ENERGY System, a DMH tool capable of screening for resource and mental health needs and providing automated, personalized, and free resources and techniques to meet the identified needs. Future studies should seek direct feedback from community members to further improve the ENERGY System and its dissemination to encourage use.

(JMIR Form Res 2022;6(6):e38162) doi:10.2196/38162

KEYWORDS

digital mental health; underserved; health disparities; COVID-19; screening; referral; mental health; digital health; feasibility study; mobile app; mHealth; mobile health; emotional need; digital health tool; health resource; health care cost

Introduction

Background

Prior to the COVID-19 pandemic, nearly half of all adolescents and a quarter of all adults in the United States met the criteria for at least one mental health disorder [1,2]. The effects of the COVID-19 pandemic have exacerbated mental health concerns, particularly for adolescents and young adults [3-5]. Further, structural barriers that existed prior to the COVID-19 pandemic primed specific communities to disproportionately experience negative effects from the pandemic [6,7]. However, the majority of these individuals have not received mental health care [8,9]. Mental health stigma, time limitations, and the identification of mental health problems were among the primary barriers to accessing care [10-20]. Additionally, the demand for behavioral health care frequently exceeds clinic capacity, particularly in rural areas [21]. Given the lack of accessible mental health services and the increased need for such resources, it is imperative to identify novel means for providing behavioral health assessment and care, particularly to members of vulnerable communities.

Digital mental health (DMH) tools are one way to enhance the accessibility of evidence-based interventions and resources [22]. DMH refers to the use of technology platforms, such as websites (ie, eHealth), mobile apps (ie, mobile health), and electronic devices (eg, Fitbit and Apple Watch), to deliver behavioral health assessments and interventions. The use of DMH has expanded rapidly during the COVID-19 pandemic [23], and there are multiple reasons it represents a viable means to provide continued behavioral health resources, particularly for communities disproportionately facing access barriers. First, DMH tools can assess and offer resources for use in real time and real-world environments without requiring individuals to travel to mental health care settings and allow for asynchronous, socially distanced care [24]. Further, DMH can also link to local or web-based resources for basic needs (eg, housing support); facilitating access to resources that support basic physiological needs is consistent with tools aimed at supporting higher level needs, such as mental health [25]. Second, mobile device access is ubiquitous [26], and the use of DMH therefore increases the likelihood of engaging individuals through technology that they are already using (eg, smartphones and tablets). Finally, youths are more likely to seek and access emotional support through technology compared to in-person care [26,27], and the pandemic has increased technology use and ownership across all ages, including older adults [28]. This makes DMH a delivery mechanism that harnesses an environment in which those at the highest risk for pandemic-related mental health concerns are more likely to feel comfortable. Importantly, DMH tools developed prior to and during the COVID-19 pandemic have the potential to extend in utility beyond the current pandemic to potential future crises [29].

The digital marketplace for DMH tools can be overwhelming with app stores featuring over 10,000 publicly available mobile health apps. The majority of available DMH tools are either not based on evidence-based interventions or have not been rigorously evaluated [28]. In addition to the overwhelming amount of available DMH tools, most require users to have already identified the problems for which they are seeking help, which can be difficult for some [30]. Thus, to make DMH more accessible to vulnerable populations, potential users must be able to identify (1) what problems they are having (eg, anxiety and depression, etc) and (2) which tools are appropriate and efficacious for their needs. Moreover, it is important that interventions be transparent and specifically targeted toward vulnerable groups and adapted to their needs.

Objective

To address these challenges, we developed a DMH tool, the Emotional Needs Evaluation and Resource Guide for You (ENERGY) System, to briefly assess mental health and other resource needs and provide automated, personalized feedback based on the identified needs. The objective of this study was to develop the tool and assess its feasibility (ie, engagement and satisfaction) to guide future developments. The ENERGY System was originally designed in the context of the COVID-19 pandemic for adolescents and young adults (AYAs) and their families residing in the West and South Side communities of Chicago but was ultimately deployed throughout the city and surrounding neighborhoods of Chicago, Illinois. These communities have faced stark mental and behavioral health disparities, both prior to and during the pandemic [7,31-33].

Methods

Participants

Participants were recruited from the Chicagoland area (ie, the city of Chicago, Illinois, and its surrounding suburbs and towns). Recruitment was conducted via (1) the distribution of flyers that contained the study link and QR code to the Rush University Education and Career Hub and local community partners and at community events (eg, Easter egg hunt and vaccine clinics); (2) the posting of flyers and short videos on social media platforms (eg, departmental Twitter account); and (3) targeted messaging to electronic health records (eg, the Epic MyChart

app) of patients receiving services from 3 school-based health centers located in Chicago's West Side communities. The initial deployment of the ENERGY System in March 2021 limited participation to people aged 15-25 years or caregivers of children aged <18 years. However, due to community needs and recognition that families with complex needs extend beyond these subgroups, the inclusion criteria were expanded at the end of May 2021 to include anyone aged ≥15 years.

Ethics Approval

The study procedures were approved by the Rush University Medical Center Institutional Review Board (20092408).

Enrollment Procedure

Waivers of the documentation of informed consent and parental permission as well as an alteration of consent were obtained. Interested individuals accessed the system via a REDCap [34] link provided on recruitment materials (eg, physical flyers, social media posts, and email blurbs); the REDCap link opened a page on the participant's personal device's web browser. REDCap was used to increase accessibility, as users would not be required to download anything to their device to use the ENERGY System. After clicking the link, participants were asked to enter their age and whether they are a parent of a child aged <18 years. Eligible persons were directed to a brief form listing all the required elements of consent. Participants were asked to click "I agree" to continue to the ENERGY System. Those who did not agree to the elements of consent were exited from the system.

ENERGY System

System Overview

The ENERGY System is a brief, DMH-screening, and automated resource identification system. The ENERGY System was developed to provide individuals with an automated tool that can identify various emotional and mental health concerns and provide an automated list of appropriate self-help and other resources to address these identified concerns. Automating the screening and resource identification process was intended to improve the accessibility of such services, which are usually performed by mental health or social service providers and thus have limited scalability. Although the ENERGY System was administered via REDCap [34] for this study, it can also be transported to other platforms.

ENERGY System Development

The ENERGY System was designed to minimize the burden related to the assessment of current needs and concerns and maximize the feasibility. To achieve this, the mental health screening portion of the ENERGY System combines 16 items drawn from existing, validated scales assessing problem-solving issues, depression and anxiety symptoms, as well as alcohol and substance use into a single survey. The items used in the ENERGY System were derived from prior research on the development of brief assessments of each of these constructs. Specifically, the items used in the ENERGY System were drawn from the 7-item Generalized Anxiety Disorder Assessment (GAD-7) [35], 5-item PTSD Checklist for DSM-5 (PCL-5) [36], 9-item Patient Health Questionnaire (PHQ-9) [37], 3-item Alcohol Use Disorder Identification Test-Consumption (AUDIT-C) [38], and Drug Abuse Screening Test (DAST-10) [39]. In a separate research study (Christ et al, unpublished data, 2022), our team used Item Response Theory on several of these short forms and the full scales from which they were derived to identify specific items that best represent the underlying latent constructs. For each brief scale derived based on Item Response Theory, the associations with the respective full scales as well as related constructs were assessed. Furthermore, our team evaluated the predictive validity of these brief scales by examining how changes in the brief scales predicted changes in the related constructs over the course of treatment and how this compared to the full scales (Christ et al, unpublished data, 2022). For the ENERGY System, we only retained items/brief scales that explained the underlying constructs and were able to adequately capture changes in these constructs following an intervention. This step was important to ensure that the ENERGY System would be able to accurately reflect changes in the underlying constructs when administered repeatedly. The resulting 16-item ENERGY System uses 2 items from the PCL-5, 4 items from the PHQ-9, 3 items from the GAD-7, 3 items from the AUDIT-C, 3 items from the DAST-10, and 1 author-developed item asking about participants' ability to handle problems.

The ENERGY System screening items were then mapped onto 6 different intervention domains. Anxiety-related symptoms were mapped to Relaxation strategies. Depressive symptoms that were primarily associated with the lack of activity were mapped to Behavioral Activation strategies. Anxiety and depressive symptoms that were primarily associated with thinking processes were mapped to Cognitive Restructuring strategies. Problem-solving was mapped to Specific, Measurable, Achievable, Relevant, and Time-bound goals and other Problem-solving strategies. Alcohol and substance use were mapped to Alcohol and Substance Use Management and Harm Reduction strategies. These domains were created to identify specific intervention targets that could be more easily matched to specific evidence-based cognitive behavioral skills. For example, although anxiety is a single construct, different interventions may be more effective depending on the symptom presentation. An individual with anxiety symptoms of feeling on edge or having difficulty relaxing may benefit more from exercises focused on activating the parasympathetic nervous system (eg, diaphragmatic breathing, progressive muscle relaxation, and mindfulness), whereas an individual who has difficulty managing their worried thoughts may benefit from cognitive restructuring. Expert-reviewed resources, including MindTools [40] and PsyberGuide [41], were used to create a repository of DMH tools that have been shown to be beneficial for depression, anxiety, alcohol use, substance use, and problem-solving within personalized resources (eg, COVID Coach App [42]).

In addition to assessing specific emotional and mental health concerns, a goal of the ENERGY System was to capture resource needs and provide immediate, relevant information about readily accessible resources specific to the location in which the ENERGY System was first tested (ie, the West and South Side communities of Chicago). Namely, participants were

directed to indicate whether they needed assistance with any of the following domains: housing, finances, parenting, childcare, food, transportation, medical care, pet care, elderly care, and any other miscellaneous areas of concern. Participants clicked a box next to any domain for which they wanted to receive assistance. These domains were selected because deficits in these areas have been associated with stress and negative impacts on individuals and their families in the short and long term [43]. Local community resources were identified through various methods. A web-based search for local Chicago resources (eg, lists of local food banks and rent/mortgage assistance programs) was conducted, and recommendations for COVID-19 testing sites and community safety hotlines were included to meet the needs of Chicago residents. Additionally, the team incorporated pertinent resources from a pre-existing list of community resources created by a team member for use during clinical care. Community resources were also identified through organizational LISTSERV software applications, colleagues, and a list of resources sent by a local congressman to his constituents. All community resources were checked for accuracy to ensure that they were still available as advertised prior to publishing them into recommendations for study participants.

ENERGY System Mental Health and Resource Feedback

Immediately upon the completion of the screening questions, the ENERGY System automatically scored the assessments to identify the domains that could benefit from intervention. Positive endorsement of an intervention domain was defined as a severity rating of at least 50% of the total possible score for the domain. For each positive domain, the ENERGY System automatically provided the individual with personalized feedback (eg, "Based on your responses, you may benefit from learning some relaxation strategies") and community-based and digital resources tailored to participant responses. Resources were presented in a person-focused rather than clinical manner (eg, using terms such as "sadness" rather than "depression") to reduce stigma.

The emotional and mental health feedback involved specific cognitive behavioral techniques that have been identified in prior research to be particularly effective for the respective concerns. For example, individuals who screened positive for sadness were introduced to behavioral activation techniques (referred to as "Increasing Activities"). Those who screened positive for symptoms of anxiety were provided tips and resources around progressive muscle relaxation, self-soothing, and meditation (referred to as "Relaxation"). All techniques were framed in a self-guided format by cognitive-behaviorally trained psychologists and described so that they could be easily followed by users without requiring coaching from a professional. In addition to detailed instructions, links to publicly available resources explaining these techniques (eg, YouTube) were provided to present individuals with added options for learning the different skills.

Feedback about other available resources was presented in a similar manner. Requested resources were presented as brief excerpts detailing the types of services provided and linking individuals to the agencies' websites where more details could be found. Each resource opened to a new tab so that individuals could keep their list of recommended resources open in their browser for as long as necessary. Individuals also had the option to enter their email address to have their list automatically emailed to them. The resources feedback page also contained the link to the ENERGY System for the individual to visit again at any time or to share with others.

Measures

Demographic Characteristics

Participants were asked to provide demographic characteristics, including age, gender, ethnicity, race, sexual orientation, the highest level of education, current student/employment status, family's total annual income before taxes, zip code, and parental status.

ENERGY System

As described above, the ENERGY System assessment consisted of 16 items drawn from previously validated symptom scales. The reliabilities of the items in each of the intervention domains were adequate to good (Relaxation: Cronbach α=.77; Behavioral Activation/Increasing Activities: Cronbach α=.83; and Cognitive Restructuring [Low Mood]: Cronbach α=.86). Cronbach α was not calculated for Problem-solving as this was assessed via a single item. Cronbach α's were also not calculated for Alcohol and Substance Use Management as items on the underlying scales were only triggered after the initial question about any alcohol or substance use was positively endorsed. Thus, participants had a widely varying number of responses to these items.

ENERGY System Satisfaction

Immediately after reviewing their recommended resources, participants were given the opportunity to provide overall feedback about the ENERGY System. Elicited feedback involved overall satisfaction and satisfaction with the ease of answering the questions, amount of time it took to complete, and recommended resources. Participants were also asked how likely they were to use the resources in the future and recommend the ENERGY System to others. Finally, participants were asked whether they would be interested in receiving professional support for resources or emotional needs, such as via connection to a therapist or primary care provider, if this were to be offered in the future.

COVID-19 Exposure and Family Impact

The COVID-19 Exposure and Family Impact Survey (CEFIS) [44] or the COVID-19 Exposure and Family Impact Survey Adolescent and Young Adult Version (CEFIS-AYA) [45] were administered, depending on age (CEFIS for all adults aged ≥30 years and CEFIS-AYA for AYAs aged 15-29 years). The CEFIS and CEFIS-AYA assess the impact of the COVID-19 pandemic on adolescents, young adults, and families. The CEFIS yields (1) an Exposure score (scores range from 0-25, with higher scores indicating greater exposure); (2) an Impact score (scores averaged from a 4-point Likert scale with scores >2.5 considered positively valenced, meaning more impact, and scores <2.5 considered negatively valenced, meaning less impact); (3) a Distress score (scores range from 1-10, with higher scores

indicating greater distress); and (4) an open-ended question to promote the sharing of details not covered. The CEFIS-AYA differs from the CEFIS by having 3 additional Exposure items, 6 additional Impact items, and a single item for the Distress scale to reflect the experience of AYAs compared to that of a family. The CEFIS demonstrated acceptable reliability for the current sample (αs>.73). However, due to the possibility that adult respondents completed the CEFIS without having children, the CEFIS Distress was reported at the item level (personal and child, separately) as opposed to the scale level (personal and child combined). The CEFIS-AYA demonstrated acceptable reliability across subscales (αs>.69).

Data Analysis

Descriptive analyses were used to (1) characterize the sample in terms of demographic characteristics and reported resource needs; (2) determine use rates; (3) identify the total provided mental health resources; and (4) assess user satisfaction with the ENERGY System.

Results

Participants

Table 1 depicts the demographic characteristics of the sample. Of the 110 participants who provided demographic data, they were primarily AYAs (68.2%, n=75), cisgender female (81.3%, n=91), heterosexual or straight (67.6%, n=75), non-Hispanic/Latinx (67.3%, n=74), and Black or African American (40.9%, n=45) or White (40%, n=44). A total of 103 (93.6%) participants provided their zip codes. Of these 103 participants, the majority (95.1%, n=98) lived in Illinois, with 80 (77.7%) residing within the city limits of Chicago. The remaining 5 (4.9%) were from California, Indiana, New York, Texas, and Wisconsin.

Table 1. Participant characteristics.

Characteristic	Adolescents and young adults[a] (n=75)	Adults[a] (n=35)	All participants (N=110)
Age (years), mean (SD; range)	18.95 (3.18; 15-25)	43.89 (13.94; 26-71)	26.88 (14.27; 15-71)
Gender, n (%)			
Cisgender female	68 (90.7)	22 (62.9)	90 (81.8)
Cisgender male	4 (5.3)	11 (31.4)	15 (13.6)
Transgender female	0 (0)	1 (2.9)	1 (0.9)
Transgender male	1 (1.3)	1 (2.9)	2 (1.8)
Nonbinary	1 (1.3)	0 (0)	1 (0.9)
Prefer not to answer	1 (1.3)	0 (0)	1 (0.9)
Ethnicity, n (%)			
Hispanic/Latinx	28 (37.3)	7 (20)	35 (31.8)
Non-Hispanic/Latinx	47 (62.7)	27 (77.1)	74 (67.3)
Prefer not to answer	0 (0)	1 (2.9)	1 (0.9)
Race[b], n (%)			
American Indian or Alaskan Native	1 (1.3)	1 (2.9)	2 (1.8)
Asian	5 (6.7)	2 (5.7)	7 (6.4)
Black or African American	38 (50.7)	7 (20)	45 (40.9)
White	20 (26.7)	24 (68.6)	44 (40)
Other	5 (6.7)	1 (2.9)	6 (5.5)
Prefer not to answer	10 (13.3)	1 (2.9)	11 (10)
Sexual orientation, n (%)			
Heterosexual or straight	51 (68)	24 (68.6)	75 (68.2)
Gay or lesbian	3 (4)	5 (14.3)	8 (7.3)
Bisexual or pansexual	18 (24)	4 (11.4)	22 (20)
Asexual	0 (0)	1 (2.9)	1 (0.9)
Other	1 (1.3)	0 (0)	1 (0.9)
Prefer not to answer	2 (2.7)	1 (2.9)	3 (2.7)
Highest level of education, n (%)			
Finished grade school	5 (6.7)	0 (0)	5 (4.5)
Some high school	29 (38.7)	0 (0)	29 (26.4)
Finished high school	9 (12)	0 (0)	9 (8.2)
Business or technical school	0 (0)	1 (2.9)	1 (0.9)
Some college	21 (28)	5 (14.3)	26 (23.6)
Finished college	8 (10.7)	8 (22.9)	16 (14.5)
Some graduate or professional school	1 (1.3)	6 (17.1)	7 (6.4)
Finished graduate or professional school	2 (2.7)	15 (42.9)	17 (15.5)

[a]Adolescents and young adults were aged 15-25 years and adults were aged ≥26 years.

[b]The total number of reported racial identities is 115 due to participants identifying with more than one racial category.

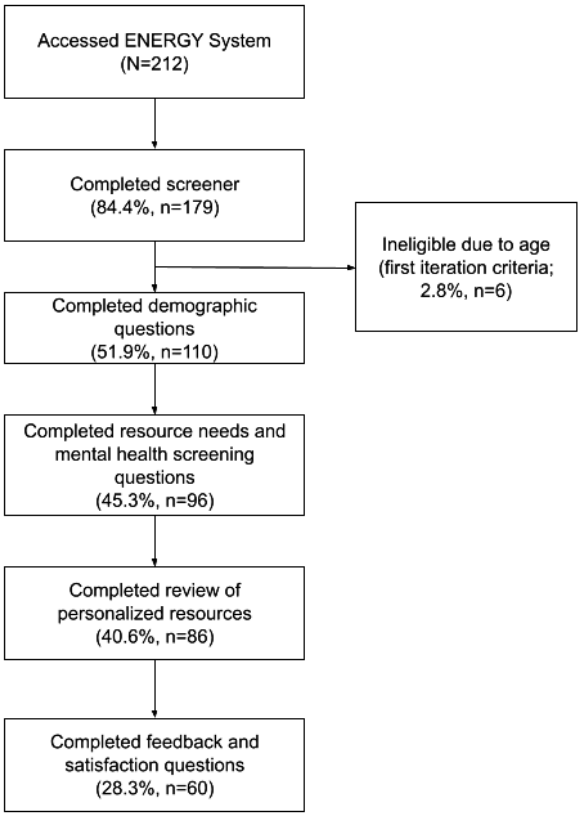
ENERGY System Use Data

Between August 2020 and December 2021, the ENERGY System link was accessed 212 times. Attrition occurred with each advancement of the system, with 84.4% (n=179) completing the screening questions, 51.9% (n=110) completing the demographic questions, 45.3% (n=96) completing the resource needs and mental health screening, 40.6% (n=86) completing the review of their personalized resources, and 28.3% (n=60) providing feedback about the ENERGY System. On average, participants completed their interactions with the system in 12 (SD 8.33) minutes (6 outliers were removed from

the average calculation as it was believed that they completed their interactions across multiple visits [ie, >1 hour], with a range from 1 hour and 20 minutes to 23 hours and 29 minutes).

Figure 1 depicts the flow chart for the ENERGY System interactions.

Figure 1. Flow chart of participants through ENERGY System use. ENERGY: Emotional Needs Evaluation and Resource Guide for You.



Mental Health Intervention Domains

A total of 96 participants (45.3%, N=212) completed the mental health screening questions. Participant responses on the mental health screening questions triggered on average 2.04 (SD 1.94) intervention domains. Of the 96 responses, Behavioral Activation/Increasing Activities was the most launched intervention domain (56%, n=54), followed by Relaxation (44%, n=42), Problem-solving (41%, n=39), and Cognitive Restructuring (Low Mood; 32%, n=31). Alcohol Use and Substance Use were the least frequently launched intervention domains, launching only 4 (4%) times each.

Resource Needs

A total of 96 participants (45.3%, N=212) completed resource needs questions, with half (50%, n=48) responding in a way that indicated that they did not need assistance with any resources. As such, the average number of resource categories provided to participants was 0.69 (SD1.44; range 0-7). Of the 96 responses, the endorsed resource needs included financial support (33%, n=32), transportation (26%, n=25), food (24%, n=23), housing (17%, n=16), medical care (14%, n=13), parenting (12%, n=12), pet care (11%, n=11), childcare (10%, n=10), and resources for caretakers of older adults (5%, n=5).

COVID-19 Exposure and Impact

A total of 64 AYA participants and 24 adult participants completed some portion of the CEFIS-AYA and CEFIS, respectively. To limit participant burden, Exposure subscale items were offered optionally and completed by 16 (25%, N=64) AYAs and 17 (71%, N=24) adults. The AYAs and adults had an average Exposure rating of 13.13 (SD 4.00; range 6-20) and 8.65 (SD 3.02; range 3-16), respectively. Both the AYA and adult Impact subscale scores were positively valenced (more impact), with AYAs having an average Impact score of 2.82 (SD 0.65; range 1.21-4.00) and adults having an average Impact score of 3.00 (SD 0.52; range 2.10-3.78). AYAs endorsed an average personal Distress rating of 5.78 (SD 2.44; range 1-10). Adults reported their personal Distress on average as 6.46 (SD 2.41; range 2-10) and their children's Distress was rated on average as 4.64 (SD 3.01; range 1-10).

ENERGY System Satisfaction

A total of 60 (28.3%, N=212) participants provided feedback about the ENERGY System following use. Of the 60 participants, the majority agreed or strongly agreed that they were satisfied with the system overall (65%, n=39), ease of answering questions (87%, n=52), and time it took to answer questions (90%, n=54). The majority (75%, n=45) also agreed or strongly agreed that the ENERGY System asked about all of their needs. Finally, the majority (78%, n=47) somewhat agreed, agreed, or strongly agreed that they were likely to use the system's recommended resources.

Discussion

Principal Findings

This study aimed to develop and assess the feasibility of a DMH tool that (1) screened for resource and mental health needs and (2) provided automated, personalized, and free resources and techniques to meet the identified needs. Dissemination efforts focused on institutional and community partner collaborations to focus on use by residents of communities facing disproportionate behavioral health disparities. Over 200 community members, largely within the Chicagoland area, demonstrated initial interest in the ENERGY System. However, attrition occurred with each progressing stage of the system (ie, screener, demographic questions, resource needs and mental health screening assessment, review of personalized resources, and system feedback). Half of all individuals who completed the screening questions endorsed not having resource needs. On average, individuals received resources for 2 mental health symptoms and lists based on their responses. As is common for subjective evaluations of DMH, even in the face of relatively low real-world engagement [46], users who provided feedback were primarily satisfied with the system.

The current sample endorsed higher than average impacts from the COVID-19 pandemic, with AYAs reporting higher exposure to pandemic experiences than their adult counterparts [44]. These experiences have occurred within the context of historical and current behavioral health disparities and broader community hardships [32]. As such, the COVID-19 pandemic has exacerbated the already existent need for scalable, accessible, and affordable DMH, particularly for vulnerable communities [47]. Although multiple studies of pandemic-specific DMH interventions have occurred [48], the ENERGY System is unique in its intended use (ie, single or repeated use) and potential for extended utility beyond the pandemic (eg, compliant with but not centered upon social distancing practices). However, to be useful for communities moving forward, multiple limitations will need to be addressed.

Accessibility

The ENERGY System was intended to serve those facing the greatest hardships and disparities from the pandemic. However, the majority of sample members denied requiring resource needs, and about 40% of the adults who used the system had a graduate or professional degree. These findings likely indicate that our recruitment efforts did not reach many of the residents requiring this additional support. Further, slightly over 200 initial clicks to the system suggest that the deployment had a low reach (eg, Chicago's Lower West Side community has a population of about 12,000 people per square mile) [49]. Therefore, increasing accessibility is an important factor for future deployments of the ENERGY System and similar DMH tools. The ENERGY System was designed for an immediate need and in the context of social distancing mandates. However, best practices would include the incorporation of human-centered design [50-52] and community-based participatory research methodologies [53] to ensure that likely end users are active collaborators—those with agency—throughout the design and deployment phases. Doing so better ensures that DMH is accessible to and appropriate for the intended users [22,54]. For example, nearly one-third of the sample identified as Hispanic or Latinx, but the tool was only available in the English language. The use of human-centered design practices with community members would provide ongoing assessment of likely end users' native languages and language preferences for DMH, promoting the likelihood of increasing accessibility to nonnative English speakers.

Engagement

Less than half of the sample members completed the resource needs and mental health screening questions. The inclusion of additional questions for the purposes of research (eg, CEFIS and satisfaction questions) likely contributed to some of the attrition that occurred with the ENERGY System in this study. We note that engagement with DMH in real-world settings is generally low [55], with top barriers including (1) stigma, (2) problem recognition, and (3) knowledge of treatment options [10-12]. The ENERGY System is designed to address these concerns (ie, anonymous participation on a personal device to address stigma; feedback about mental health symptoms to address problem recognition; and free resources provided to address knowledge of treatment options). However, engagement with the system was low. These findings again support the need to involve representative potential users throughout the design process. It is unsurprising that community members systemically experiencing disproportionate disparities may not engage well with DMH in real-world settings, as DMH has historically been developed without their input regarding their lived experiences and needs [56]. A focus on engagement strategies with systemically excluded and marginalized community members must also be a focus of future design and research.

Limitations

This study should be considered in light of its specific limitations. First, social distancing mandates limited recruitment efforts. Namely, typical recruitment methods of community partners often involve in-person communication (even in the case of sharing a website link or QR code), which was limited due to the pandemic. A large portion of flier dissemination occurred through mass emails; this methodology likely resulted in study materials being overshadowed by other web-based obligations [57,58]. Second, remote learning or e-learning limited the ability of recruitment partners in school settings to engage and explain the study to younger potential participants. Third, the ENERGY System was only provided in the English language. Roughly 16% of Chicago residents do not speak English as a primary language, with Spanish being the most frequently spoken language among this group [59]. This barrier left several Chicago families unable to complete the screener. Future resource tools similar to the ENERGY System should create multiple translations, based on formative assessments of preferred language, for optimal use among diverse communities. Finally, the ENERGY System was disseminated through the Rush University Medical Center clinics and community partnerships. Although some users may trust DMH when overseen by a university setting due to institutional review board oversight, others are less inclined to trust university and larger health systems–sponsored DMH due to historical injustices

[60]. Future research should collaborate with community members to assess how the role of academic institution involvement may promote or hinder DMH engagement.

Conclusions

This study provided initial support for the feasibility of the ENERGY System, a DMH tool capable of screening for resource and mental health needs and providing automated, personalized, and free resources and techniques to meet the identified needs. Future research should involve collaboration with community members and collect more detailed formative and summative data from representative end users of the ENERGY System. As such, better attempts to meet the needs of users with intersectional identities and varying mental health and resource needs may result in improved engagement [46]. Additionally, larger scale research is needed to determine the feasibility of the ENERGY System beyond the areas targeted in this study. Moreover, longitudinal studies are needed to determine whether individuals use the resources and mental health skills they were provided and whether the use of skills and resources improves their reported symptoms over time.

Acknowledgments

This research was supported in part by an R3 funding mechanism supported by the Rush University Medical Center and RTI International. CSS was also supported by fellowships from the National Institute of Mental Health (K08 MH125069) and the Cohn Family Foundation. PH was supported by a career development award from the National Center for Advancing Translational Sciences (KL2-TR002387) and received additional grant from the Wounded Warrior Project. We thank the Rush Community-Based Practices, comprising the school-based health centers and the Adolescent Family Center; Rush Education and Career Hub; and our West Side community partners, without whom this research would not have been possible.

Conflicts of Interest

None declared.

References

1. Merikangas KR, He JP, Burstein M, Swanson SA, Avenevoli S, Cui L, et al. Lifetime prevalence of mental disorders in U.S. adolescents: results from the National Comorbidity Survey Replication--Adolescent Supplement (NCS-A). J Am Acad Child Adolesc Psychiatry 2010 Oct;49(10):980-989. [doi: 10.1016/j.jaac.2010.05.017] [Medline: 20855043]
2. Key substance use and mental health indicators in the United States: results from the 2020 National Survey on Drug Use and Health. Substance Abuse and Mental Health Services Administration. 2021 Oct. URL: https://www.samhsa.gov/data/sites/default/files/reports/rpt35325/NSDUHFFRPDFWHTMLFiles2020/2020NSDUHFFR1PDFW102121.pdf [accessed 2022-06-14]
3. Magson NR, Freeman JYA, Rapee RM, Richardson CE, Oar EL, Fardouly J. Risk and protective factors for prospective changes in adolescent mental health during the COVID-19 pandemic. J Youth Adolesc 2021 Jan;50(1):44-57. [doi: 10.1007/s10964-020-01332-9] [Medline: 33108542]
4. Czeisler, Lane RI, Wiley JF, Czeisler CA, Howard ME, Rajaratnam SMW. Follow-up survey of US adult reports of mental health, substance use, and suicidal ideation during the COVID-19 pandemic, September 2020. JAMA Netw Open 2021 Feb 01;4(2):e2037665 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.37665] [Medline: 33606030]
5. Chadi N, Spinoso-Di Piano C, Osmanlliu E, Gravel J, Drouin O. Mental health-related emergency department visits in adolescents before and during the COVID-19 pandemic: a multicentric retrospective study. J Adolesc Health 2021 Nov;69(5):847-850 [FREE Full text] [doi: 10.1016/j.jadohealth.2021.07.036] [Medline: 34462192]
6. Dorn AV, Cooney RE, Sabin ML. COVID-19 exacerbating inequalities in the US. Lancet 2020 Apr 18;395(10232):1243-1244. [doi: 10.1016/S0140-6736(20)30893-X] [Medline: 32305087]
7. Pierce JB, Harrington K, McCabe ME, Petito LC, Kershaw KN, Pool LR, et al. Racial/ethnic minority and neighborhood disadvantage leads to disproportionate mortality burden and years of potential life lost due to COVID-19 in Chicago, Illinois. Health Place 2021 Mar;68:102540 [FREE Full text] [doi: 10.1016/j.healthplace.2021.102540] [Medline: 33647635]
8. Walker ER, Cummings JR, Hockenberry JM, Druss BG. Insurance status, use of mental health services, and unmet need for mental health care in the United States. Psychiatr Serv 2015 Jun;66(6):578-584. [doi: 10.1176/appi.ps.201400248] [Medline: 25726980]
9. Merikangas KR, He J, Burstein M, Swendsen J, Avenevoli S, Case B, et al. Service utilization for lifetime mental disorders in U.S. adolescents: results of the National Comorbidity Survey-Adolescent Supplement (NCS-A). J Am Acad Child Adolesc Psychiatry 2011 Jan;50(1):32-45. [doi: 10.1016/j.jaac.2010.10.006] [Medline: 21156268]
10. Radez J, Reardon T, Creswell C, Lawrence PJ, Evdoka-Burton G, Waite P. Why do children and adolescents (not) seek and access professional help for their mental health problems? a systematic review of quantitative and qualitative studies. Eur Child Adolesc Psychiatry 2021 Feb 21;30(2):183-211 [FREE Full text] [doi: 10.1007/s00787-019-01469-4] [Medline: 31965309]

11. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psychiatry 2010 Dec 30;10:113 [FREE Full text] [doi: 10.1186/1471-244X-10-113] [Medline: 21192795]
12. Wisdom JP, Clarke GN, Green CA. What teens want: barriers to seeking care for depression. Adm Policy Ment Health 2006 Mar;33(2):133-145. [doi: 10.1007/s10488-006-0036-4] [Medline: 16489480]
13. Reardon T, Harvey K, Baranowska M, O'Brien D, Smith L, Creswell C. What do parents perceive are the barriers and facilitators to accessing psychological treatment for mental health problems in children and adolescents? a systematic review of qualitative and quantitative studies. Eur Child Adolesc Psychiatry 2017 Jun 4;26(6):623-647. [doi: 10.1007/s00787-016-0930-6] [Medline: 28054223]
14. Hyde PS, Substance Abuse and Mental Health Services Administration. Report to Congress on the nation's substance abuse and mental health workforce issues. California Institute for Behavioral Health Solutions. 2013 Jan 24. URL: http://www.cimh.org/sites/main/files/file-attachments/samhsa_bhwork_0.pdf [accessed 2022-06-14]
15. Salloum A, Johnco C, Lewin AB, McBride NM, Storch EA. Barriers to access and participation in community mental health treatment for anxious children. J Affect Disord 2016 May 15;196:54-61. [doi: 10.1016/j.jad.2016.02.026] [Medline: 26901657]
16. Douthit N, Kiv S, Dwolatzky T, Biswas S. Exposing some important barriers to health care access in the rural USA. Public Health 2015 Jun;129(6):611-620. [doi: 10.1016/j.puhe.2015.04.001] [Medline: 26025176]
17. Allen EM, Call KT, Beebe TJ, McAlpine DD, Johnson PJ. Barriers to care and health care utilization among the publicly insured. Med Care 2017 Mar;55(3):207-214. [doi: 10.1097/MLR.0000000000000644] [Medline: 27579910]
18. Hacker K, Anies M, Folb BL, Zallman L. Barriers to health care for undocumented immigrants: a literature review. Risk Manag Healthc Policy 2015;8:175-183 [FREE Full text] [doi: 10.2147/RMHP.S70173] [Medline: 26586971]
19. Stevens J, Kelleher KJ, Ward-Estes J, Hayes J. Perceived barriers to treatment and psychotherapy attendance in child community mental health centers. Community Ment Health J 2006 Oct;42(5):449-458. [doi: 10.1007/s10597-006-9048-5] [Medline: 16964565]
20. de Haan AM, Boon AE, de Jong JTVM, Hoeve M, Vermeiren RRJM. A meta-analytic review on treatment dropout in child and adolescent outpatient mental health care. Clin Psychol Rev 2013 Jul;33(5):698-711. [doi: 10.1016/j.cpr.2013.04.005] [Medline: 23742782]
21. Andrilla CHA, Patterson DG, Garberson LA, Coulthard C, Larson EH. Geographic variation in the supply of selected behavioral health providers. Am J Prev Med 2018 Jun;54(6 Suppl 3):S199-S207 [FREE Full text] [doi: 10.1016/j.amepre.2018.01.004] [Medline: 29779543]
22. Lattie EG, Stiles-Shields C, Graham AK. An overview of and recommendations for more accessible digital mental health services. Nat Rev Psychol 2022 Jan 26;1(2):87-100. [doi: 10.1038/s44159-021-00003-1]
23. Clipper B. The influence of the COVID-19 pandemic on technology: adoption in health care. Nurse Lead 2020 Oct;18(5):500-503. [doi: 10.1016/j.mnl.2020.06.008] [Medline: 32837346]
24. Torous J, Jän Myrick K, Rauseo-Ricupero N, Firth J. Digital mental health and COVID-19: using technology today to accelerate the curve on access and quality tomorrow. JMIR Ment Health 2020 Mar 26;7(3):e18848 [FREE Full text] [doi: 10.2196/18848] [Medline: 32213476]
25. SAMHSA's working definition of recovery. Substance Abuse and Mental Health Services Administration. 2012. URL: https://store.samhsa.gov/sites/default/files/d7/priv/pep12-recdef.pdf [accessed 2022-06-14]
26. Anderson M, Jiang J. Teens' social media habits and experiences. Pew Research Center. 2018 Nov 28. URL: https://www.pewresearch.org/internet/2018/11/28/teens-social-media-habits-and-experiences/ [accessed 2022-06-14]
27. Anderson M, Jiang J. Pew Research Center. 2018 May 31. URL: https://www.pewinternet.org/2018/05/31/teens-social-media-technology-2018/ [accessed 2022-06-14]
28. Faverio M. Share of those 65 and older who are tech users has grown in the past decade. Pew Research Center. 2022 Jan 13. URL: https://www.pewresearch.org/fact-tank/2022/01/13/share-of-those-65-and-older-who-are-tech-users-has-grown-in-the-past-decade/ [accessed 2022-02-08]
29. Stiles-Shields C, Plevinsky JM, Psihogios AM, Holmbeck GN. Considerations and future directions for conducting clinical research with pediatric populations during the COVID-19 pandemic. J Pediatr Psychol 2020 Aug 01;45(7):720-724. [doi: 10.1093/jpepsy/jsaa055] [Medline: 32676670]
30. Evans-Lacko S, Stolzenburg S, Gronholm PC, Ribeiro W, York-Smith M, Schomerus G. Psychometric validation of the self-identification of having a mental illness (SELF-I) scale and the relationship with stigma and help-seeking among young people. Soc Psychiatry Psychiatr Epidemiol 2019 Jan 4;54(1):59-67 [FREE Full text] [doi: 10.1007/s00127-018-1602-2] [Medline: 30284599]
31. Hunt BR, Tran G, Whitman S. Life expectancy varies in local communities in Chicago: racial and spatial disparities and correlates. J Racial Ethn Health Disparities 2015 Dec;2(4):425-433. [doi: 10.1007/s40615-015-0089-8] [Medline: 26863550]
32. Lange-Maia BS, De Maio F, Avery EF, Lynch EB, Laflamme EM, Ansell DA, et al. Association of community-level inequities and premature mortality: Chicago, 2011-2015. J Epidemiol Community Health 2018 Dec;72(12):1099-1103. [doi: 10.1136/jech-2018-210916] [Medline: 30171083]

33. Dircksen JC, Prachand NG, Adams D, Bocskay K, Brown J, Cibulskis A, et al. Healthy Chicago 2.0: partnering to improve health equity. City of Chicago. 2016 Mar. URL: https://www.chicago.gov/content/dam/city/depts/cdph/CDPH/HC2.0Plan_3252016.pdf [accessed 2022-06-14]
34. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
35. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
36. Weathers FW, Litz BT, Keane TM, Palmieri PA, Marx BP, Schnurr PP. The PTSD checklist for DSM-5 (PCL-5). National Center for PTSD. 2013. URL: https://www.ptsd.va.gov/professional/assessment/adult-sr/ptsd-checklist.asp [accessed 2022-06-14]
37. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
38. Bradley KA, DeBenedetti AF, Volk RJ, Williams EC, Frank D, Kivlahan DR. AUDIT-C as a brief screen for alcohol misuse in primary care. Alcohol Clin Exp Res 2007 Jul;31(7):1208-1217. [doi: 10.1111/j.1530-0277.2007.00403.x] [Medline: 17451397]
39. Villalobos-Gallegos L, Pérez-López A, Mendoza-Hassey R, Graue-Moreno J, Marín-Navarrete R. Psychometric and diagnostic properties of the Drug Abuse Screening Test (DAST): comparing the DAST-20 vs. the DAST-10. Salud Ment 2015 Apr 14;38(2):89-94. [doi: 10.17711/sm.0185-3325.2015.012]
40. MindTools.io – find top mental health apps based on expert reviews. MindTools. URL: https://mindtools.io/ [accessed 2022-02-09]
41. Neary M, Schueller SM. State of the field of mental health apps. Cogn Behav Pract 2018 Nov;25(4):531-537. [doi: 10.1016/j.cbpra.2018.01.002] [Medline: 33100810]
42. Mobile apps: COVID Coach. PTSD: National center for PTSD. URL: https://www.ptsd.va.gov/appvid/mobile/COVID_coach_app.asp [accessed 2022-05-31]
43. Francis L, DePriest K, Wilson M, Gross D. Child poverty, toxic stress, and social determinants of health: screening and care coordination. Online J Issues Nurs 2018 Sep;23(3):2. [doi: 10.3912/OJIN.Vol23No03Man02] [Medline: 31427855]
44. Kazak AE, Alderfer M, Enlow PT, Lewis AM, Vega G, Barakat L, et al. COVID-19 exposure and family impact scales: factor structure and initial psychometrics. J Pediatr Psychol 2021 Jun 03;46(5):504-513. [doi: 10.1093/jpepsy/jsab026] [Medline: 33749794]
45. Schwartz LA, Lewis AM, Alderfer MA, Vega G, Barakat LP, King-Dowling S, et al. COVID-19 exposure and family impact scales for adolescents and young adults. J Pediatr Psychol 2022 Jun 07;47(6):631-640. [doi: 10.1093/jpepsy/jsac036] [Medline: 35459946]
46. Ng MM, Firth J, Minen M, Torous J. User engagement in mental health apps: a review of measurement, reporting, and validity. Psychiatr Serv 2019 Jul 01;70(7):538-544. [doi: 10.1176/appi.ps.201800519] [Medline: 30914003]
47. Figueroa CA, Aguilera A. The need for a mental health technology revolution in the COVID-19 pandemic. Front Psychiatry 2020;11:523 [FREE Full text] [doi: 10.3389/fpsyt.2020.00523] [Medline: 32581891]
48. Aguilera A, Hernandez-Ramos R, Haro-Ramos AY, Boone CE, Luo TC, Xu J, et al. A text messaging intervention (StayWell at Home) to counteract depression and anxiety during COVID-19 social distancing: pre-post study. JMIR Ment Health 2021 Nov 01;8(11):e25298 [FREE Full text] [doi: 10.2196/25298] [Medline: 34543230]
49. Maciag M. Population density for U.S. cities statistics. Governing. 2013 Oct 01. URL: http://www.governing.com/gov-data/population-density-land-area-cities-map.html [accessed 2022-01-07]
50. ISO 9241-210:2019 ergonomics of human-system interaction — part 210: human-centred design for interactive systems. International Organization for Standardization. 2019 Jul. URL: https://www.iso.org/standard/77520.html [accessed 2022-06-14]
51. IDEO. The Field Guide to Human-Centered Design. San Francisco, CA: IDEO; 2015.
52. An introduction to design thinking: process guide. Institute of Design at Stanford. 2010. URL: https://web.stanford.edu/~mshanks/MichaelShanks/files/509554.pdf [accessed 2022-06-14]
53. Collins SE, Clifasefi SL, Stanton J, The Leap Advisory Board, Straits KJE, Gil-Kashiwabara E, et al. Community-based participatory research (CBPR): towards equitable involvement of community in psychology research. Am Psychol 2018 Oct;73(7):884-898. [doi: 10.1037/amp0000167] [Medline: 29355352]
54. Stiles-Shields C, Potthoff LM, Bounds DT, Burns MTS, Draxler JM, Otwell CH, et al. Harnessing phones to target pediatric populations with socially complex needs: systematic review. JMIR Pediatr Parent 2020 Aug 26;3(2):e19269 [FREE Full text] [doi: 10.2196/19269] [Medline: 32845244]
55. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018 Aug;21(3):116-119. [doi: 10.1136/eb-2018-102891] [Medline: 29871870]

56. Hill C, Molitor M, Ortiz C. Racism and inequity are products of design. they can be redesigned. Medium. 2016 Nov 15. URL: https://medium.com/@equityXdesign/racism-and-inequity-are-products-of-design-they-can-be-redesigned-12188363cc6a [accessed 2022-06-14]
57. Hayes S, Priestley JL, Ishmakhametov N, Ray HE. "I'm not working from home, I'm living at work": perceived stress and work-related burnout before and during COVID-19. PsyArXiv Preprint posted online on July 8, 2020. [doi: 10.31234/osf.io/vnkwa]
58. Morgantini LA, Naha U, Wang H, Francavilla S, Acar, Flores JM, et al. Factors contributing to healthcare professional burnout during the COVID-19 pandemic: a rapid turnaround global survey. PLoS One 2020;15(9):e0238217 [FREE Full text] [doi: 10.1371/journal.pone.0238217] [Medline: 32881887]
59. Language access. City of Chicago. URL: https://www.chicago.gov/city/en/depts/mayor/supp_info/office-of-new-americans/language-access.html [accessed 2022-01-07]
60. Lattie EG, Stiles-Shields C, Graham AK. An overview of and recommendations for more accessible digital mental health services. Nat Rev Psychol 2022 Jan 26;1(2):87-100. [doi: 10.1038/s44159-021-00003-1]

Abbreviations

AUDIT-C: 3-item Alcohol Use Disorder Identification Test-Consumption
AYA: adolescent and young adult
CEFIS: COVID-19 Exposure and Family Impact Survey
CEFIS-AYA: COVID-19 Exposure and Family Impact Survey Adolescent and Young Adult Version
DAST-10: Drug Abuse Screening Test
DMH: digital mental health
ENERGY: Emotional Needs Evaluation and Resource Guide for You
GAD-7: 7-item Generalized Anxiety Disorder Assessment
PCL-5: 5-item PTSD Checklist for DSM-5
PHQ-9: 9-item Patient Health Questionnaire

Edited by A Mavragani; submitted 21.03.22; peer-reviewed by L Francis, A Haro-Ramos; comments to author 23.05.22; revised version received 03.06.22; accepted 05.06.22; published 22.06.22.

Please cite as:
Stiles-Shields C, Batts KR, Reyes KM, Archer J, Crosby S, Draxler JM, Lennan N, Held P
Digital Screening and Automated Resource Identification System to Address COVID-19–Related Behavioral Health Disparities: Feasibility Study
JMIR Form Res 2022;6(6):e38162
URL: https://formative.jmir.org/2022/6/e38162
doi:10.2196/38162
PMID:35696607

Letter to the Editor

Additional Measurement Approaches for Sleep Disturbances. Comment on "A Transdiagnostic Self-management Web-Based App for Sleep Disturbance in Adolescents and Young Adults: Feasibility and Acceptability Study"

Wan-Tong Tsai[1]; Tzung-Liang Liu[1]

Chung Shan Medical University, Taichung City, Taiwan

Corresponding Author:
Wan-Tong Tsai
Chung Shan Medical University
No 110, Sec 1, Jianguo N Rd
South District
Taichung City, 40201
Taiwan
Phone: 886 968938360
Email: science.tsai@gmail.com

Related Articles:

Comment on: http://www.jmir.org/2021/11/e25392/

Comment in: http://mhealth.jmir.org/2022/6/e39198/

(JMIR Form Res 2022;6(6):e35959) doi:10.2196/35959

KEYWORDS

youth; sleep; technology; mHealth; self-management; adolescents; young adults; mobile phone; smartphone; polysomnography

Carmona et al [1] analyzed the effect of DOZE (Delivering Online Zzz's with Empirical Support) on adolescents and young adults. However, this was done only via feedback from the participants. To make measurements more quantitative and less subjective, we propose a few scientific measurement approaches.

Polysomnography [2], a method to measure one's physiological state during sleep, is a common way to assess sleep quality. With the following devices, 7 parts of the physiological state can be detected:

1. Electroencephalogram: Multiple brain electrode patches are placed on the scalp to record the various stages of sleep. Thus, we can obtain a brainwave diagram for analysis.
2. Electromyography: 4 electrode patches are used to monitor muscle tension. Muscle tension will decrease significantly during sleep. We can also detect whether periodic limb movement disorder occurs and then compare the results with other data.
3. Electrocardiography: this can record the activity of the heart, including T waves, P waves, and QRS waves.
4. Electrooculography: this can record the potential difference between the cornea and the retina, so that we know when the rapid eye movement period occurs, allowing us to judge the stages of sleep.
5. Oxygen saturation: by recording this, we can know if the body remains in good condition during sleep.
6. Thoracic abdominal effort.
7. Nasal and oral airflow.

In addition, a serology examination can also be applied to quantify the survey. For instance, we can detect the melatonin levels of participants, so as to detect fluctuations from baseline to the end point. Thus, we can see if the concentration of melatonin reaches appropriate levels after participants use a web-based self-management app [3].

To keep the self-management app easy to use and convenient, any approach that requires specific tests in a hospital should not be used. Regardless, there are some take-home tests that can be administered easily. Examples include self-adhesive electrodes; heart rate monitoring equipment; or even techniques like the Freestyle Libre Pro for glucose monitoring [4], which can be used to track the physiological state of each participant.

In summary, quantifying the influence of DOZE with physiological states could make it more credible, and monitoring tests can be put into practice feasibly.

Conflicts of Interest

None declared.

References

1. Carmona NE, Usyatynsky A, Kutana S, Corkum P, Henderson J, McShane K, et al. A transdiagnostic self-management web-based app for sleep disturbance in adolescents and young adults: Feasibility and acceptability study. JMIR Form Res 2021 Nov 01;5(11):e25392 [FREE Full text] [doi: 10.2196/25392] [Medline: 34723820]
2. Jafari B, Mohsenin V. Polysomnography. Clin Chest Med 2010 Jun;31(2):287-297. [doi: 10.1016/j.ccm.2010.02.005] [Medline: 20488287]
3. Hsing AW, Meyer TE, Niwa S, Quraishi SM, Chu LW. Measuring serum melatonin in epidemiologic studies. Cancer Epidemiol Biomarkers Prev 2010 Apr;19(4):932-937 [FREE Full text] [doi: 10.1158/1055-9965.EPI-10-0004] [Medline: 20332275]
4. Blum A. Freestyle Libre glucose monitoring system. Clin Diabetes 2018 Apr;36(2):203-204 [FREE Full text] [doi: 10.2337/cd17-0130] [Medline: 29686463]

Abbreviations

DOZE: Delivering Online Zzz's with Empirical Support

Edited by T Leung; submitted 24.12.21; this is a non–peer-reviewed article;accepted 05.05.22; published 13.06.22.

Please cite as:
Tsai WT, Liu TL
Additional Measurement Approaches for Sleep Disturbances. Comment on "A Transdiagnostic Self-management Web-Based App for Sleep Disturbance in Adolescents and Young Adults: Feasibility and Acceptability Study"
JMIR Form Res 2022;6(6):e35959
URL: https://formative.jmir.org/2022/6/e35959
doi:10.2196/35959
PMID:35700003

Letter to the Editor

Authors' Response to: Additional Measurement Approaches for Sleep Disturbances. Comment on "Transdiagnostic Self-management Web-Based App for Sleep Disturbance in Adolescents and Young Adults: Feasibility and Acceptability Study"

Colleen E Carney[1], PhD; Nicole E Carmona[1], MA

Toronto Metropolitan University, Toronto, ON, Canada

Corresponding Author:
Colleen E Carney, PhD
Toronto Metropolitan University
Jorgenson Hall
350 Victoria Street
Toronto, ON, M5B 2K3
Canada
Phone: 1 4169795000 ext 552177
Email: ccarney@ryerson.ca

Related Articles:

Comment on: http://www.jmir.org/2022/6/e35959/

Comment on: http://www.jmir.org/2021/11/e25392/

(JMIR Form Res 2022;6(6):e39198) doi:10.2196/39198

KEYWORDS

youth; sleep; technology; mHealth; self-management; adolescents; young adults; mobile phone; smartphone; polysomnography

We thank Tsai and Liu [1] for their letter on our paper [2]. It is an interesting suggestion to include secondary measures of differing "objective" sleep constructs, such as electrical activity in a polysomnography (PSG), serum melatonin, glucose, or movement; however, prospective monitoring of habits and perception is the key measured construct in sleep health, as well as the diagnosis and treatment of insomnia. There was an elevated rate of insomnia symptoms in the sample, and it is important to remember that insomnia often does not have objective disturbance [3], as it is a disorder of self-reported complaint and experience. This is one of the reasons why PSG is not indicated for insomnia except to rule out other sleep disorders; other reasons include the fact that a single-night assessment is not useful, and in-lab testing can lead to dubious results because of conditioned arousal [4]. Data from the PSG is not used for an insomnia diagnosis nor is it used to determine if a treatment is effective; the patient does that.

To say that the study should be more quantitative negates that prospective sleep logging is both quantitative and the consensus tool for the field [5]. What matters in insomnia is the person's perspective. Once we move into measurement of sleep via electrical activity, chemical activity, or movement, we are one step removed from the construct of interest—the person's experience of sleeplessness. The idea to measure melatonin is somewhat curious, given that there is no reason to expect any melatonin issues to measure, or to detect an improvement in melatonin, unless there was a specific circadian rhythm disorder in the participant.

This seems a pertinent time to remind the authors of the letter [1] that these are nonclinical participants who want to improve their sleep with a self-management app. Sleep health is important for teenagers, and one would not expect anomalies in melatonin or the PSG in nonclinical users, nor would one be able to detect a change with such a restricted range (ie, these values are likely in the normal range). This is a self-management app for those with various sleep problems even in the subclinical range, and, therefore, the only aspect we could reasonably affect is the reason they would have used the app: to improve their perception of sleep. If a future researcher wanted to recruit a clinical sample, such as those with delayed sleep phase syndrome and complete assays (eg, dim light melatonin onset) or actigraphs, the app may be helpful in this regard, but given that it is self-management, and users can opt to ignore advice, it is unclear whether it would be effective in a clinical sample—again, this is for self-management purposes. Whereas some value an "objective" measure over self-report, it is

important to consider the desired construct, the use of such data, and for whom the data applies: in this case, those who self-identify as wanting better sleep and are willing to track and change their sleep habits to improve their sleep experience. Our app, designed by experts and teenage users, was successful in this regard.

Conflicts of Interest

None declared.

References

1. Tsai WT, Liu TL. Additional measurement approaches for sleep disturbances. Comment on "A transdiagnostic self-management web-based app for sleep disturbance in adolescents and young adults: Feasibility and acceptability study". JMIR Form Res 2022;6(6):e35959. [doi: 10.2196/35959]
2. Carmona NE, Usyatynsky A, Kutana S, Corkum P, Henderson J, McShane K, et al. A transdiagnostic self-management web-based app for sleep disturbance in adolescents and young adults: Feasibility and acceptability study. JMIR Form Res 2021 Nov 01;5(11):e25392 [FREE Full text] [doi: 10.2196/25392] [Medline: 34723820]
3. Edinger JD, Fins AI, Glenn DM, Sullivan JRJ, Bastian LA, Marsh GR, et al. Insomnia and the eye of the beholder: Are there clinical markers of objective sleep disturbances among adults with and without insomnia complaints? J Consult Clin Psychol 2000;68(4):586-593. [doi: 10.1037/0022-006X.68.4.586]
4. Littner M, Hirshkowitz M, Kramer M, Kapen S, Anderson WM, Bailey D, American Academy of Sleep Medicine, Standards of Practice Committe. Practice parameters for using polysomnography to evaluate insomnia: an update. Sleep 2003 Sep;26(6):754-760. [doi: 10.1093/sleep/26.6.754] [Medline: 14572131]
5. Carney CE, Buysse DJ, Ancoli-Israel S, Edinger JD, Krystal AD, Lichstein KL, et al. The consensus sleep diary: standardizing prospective sleep self-monitoring. Sleep 2012 Feb 01;35(2):287-302 [FREE Full text] [doi: 10.5665/sleep.1642] [Medline: 22294820]

Abbreviations

PSG: polysomnography

Edited by T Leung; submitted 02.05.22; this is a non–peer-reviewed article;accepted 05.05.22; published 13.06.22.

Please cite as:
Carney CE, Carmona NE
Authors' Response to: Additional Measurement Approaches for Sleep Disturbances. Comment on "Transdiagnostic Self-management Web-Based App for Sleep Disturbance in Adolescents and Young Adults: Feasibility and Acceptability Study"
JMIR Form Res 2022;6(6):e39198
URL: https://formative.jmir.org/2022/6/e39198
doi:10.2196/39198
PMID:35699990

Original Paper

Pretrained Transformer Language Models Versus Pretrained Word Embeddings for the Detection of Accurate Health Information on Arabic Social Media: Comparative Study

Yahya Albalawi[1,2,3], MSc; Nikola S Nikolov[1], PhD; Jim Buckley[1,3], PhD

[1]Department of Computer Science and Information Systems, University of Limerick, Limerick, Ireland

[2]Department of Computer and Information Sciences, College of Arts and Science, University of Taibah, Al-Ula, Saudi Arabia

[3]The Irish Software Research Centre, Lero, University of Limerick, Limerick, Ireland

Corresponding Author:
Yahya Albalawi, MSc
Department of Computer Science and Information Systems
University of Limerick
Tierney Building
Limerick, V94 T9PX
Ireland
Phone: 353 61213028 ext 3724
Email: yahalbalawi@gmail.com

Abstract

Background: In recent years, social media has become a major channel for health-related information in Saudi Arabia. Prior health informatics studies have suggested that a large proportion of health-related posts on social media are inaccurate. Given the subject matter and the scale of dissemination of such information, it is important to be able to automatically discriminate between accurate and inaccurate health-related posts in Arabic.

Objective: The first aim of this study is to generate a data set of generic health-related tweets in Arabic, labeled as either accurate or inaccurate health information. The second aim is to leverage this data set to train a state-of-the-art deep learning model for detecting the accuracy of health-related tweets in Arabic. In particular, this study aims to train and compare the performance of multiple deep learning models that use pretrained word embeddings and transformer language models.

Methods: We used 900 health-related tweets from a previously published data set extracted between July 15, 2019, and August 31, 2019. Furthermore, we applied a pretrained model to extract an additional 900 health-related tweets from a second data set collected specifically for this study between March 1, 2019, and April 15, 2019. The 1800 tweets were labeled by 2 physicians as *accurate*, *inaccurate*, or *unsure*. The physicians agreed on 43.3% (779/1800) of tweets, which were thus labeled as *accurate* or *inaccurate*. A total of 9 variations of the pretrained transformer language models were then trained and validated on 79.9% (623/779 tweets) of the data set and tested on 20% (156/779 tweets) of the data set. For comparison, we also trained a bidirectional long short-term memory model with 7 different pretrained word embeddings as the input layer on the same data set. The models were compared in terms of their accuracy, precision, recall, F_1 score, and macroaverage of the F_1 score.

Results: We constructed a data set of labeled tweets, 38% (296/779) of which were labeled as inaccurate health information, and 62% (483/779) of which were labeled as accurate health information. We suggest that this was highly efficacious as we did not include any tweets in which the physician annotators were unsure or in disagreement. Among the investigated deep learning models, the Transformer-based Model for Arabic Language Understanding version 0.2 (AraBERTv0.2)-large model was the most accurate, with an F_1 score of 87%, followed by AraBERT version 2–large and AraBERTv0.2-base.

Conclusions: Our results indicate that the pretrained language model AraBERTv0.2 is the best model for classifying tweets as carrying either inaccurate or accurate health information. Future studies should consider applying ensemble learning to combine the best models as it may produce better results.

(JMIR Form Res 2022;6(6):e34834) doi:10.2196/34834

KEYWORDS

social media; machine learning; pretrained language models; bidirectional encoder representations from transformers; BERT; deep learning; health information; infodemiology; tweets; language model; health informatics; misinformation

Introduction

Background

In the past 2 decades, there has been a dramatic increase in the number of people who use social media (SM) to participate in discussions on various topics, such as politics [1], health [2], and education [3]. Regarding health-related information, several recent studies from Saudi Arabia found that Twitter is the preferred SM platform for communicating and accessing medical information. For example, it was preferred by orthopedic surgeons to reply to (personal and professional) medical questions [4], by dental practitioners for medical consultations [5], by patients with diabetes to search for health information [6], by female students at a university in Saudi Arabia to read about systemic *lupus erythematosus* [7], and by adolescents to search for oral health information [2].

A significant problem with this form of communication is that there is no quality control over the medium, and most of the health information presented on Twitter seems inaccurate, as illustrated by the various studies summarized in Table 1. Indeed, multiple data science studies have used data sets of health-related communication on SM to study this phenomenon, and some studies [8-10] went further to design frameworks for detecting the accuracy of health information on SM.

Table 1. Summary of studies that analyzed the accuracy of health information on social media.

Studies	Number of tweets or documents	Sources	Methods to label	Language covered	Percentage of the accuracy	Topics covered	Type of study
Swetland et al [11]	358	Twitter	Expert votes; relabeling in cases of disagreement	English	25.4% inaccurate	COVID-19	Exploratory
Albalawi et al [12]	109	Twitter	Two physicians; delete if there is a disagreement	Arabic	31% inaccurate	General	Quantitative pilot study
Saeed et al [8]	208	Twitter	Expert votes; relabeling in cases of disagreement	Arabic	38% inaccurate	Cancer	ML[a]
Sharma et al [13]	183	Facebook	Two physicians; delete if there is a disagreement	English	12% inaccurate	Zika	Quantitative
Alnemer et al [14]	625	Twitter	Vote if the experts do not agree	Arabic	50% inaccurate	Only tweets from health professionals	Quantitative and exploratory study
Zhao et al [10]	5000	Health forum	Annotator voting; in addition, consulted an expert to validate information labeled as misleading	Chinese	11.4% misinformation	Autism	ML
Sell et al [15]	2460	Twitter	Coders checked the interagreement on 200 tweets	English	10% inaccurate	Ebola	Quantitative
Chew and Eysenbach [16]	5395	Twitter	Coder checked agreement on 125 tweets; unsubstantiated by the following reference standards: the CDC[b] and Public Health Agency of Canada for scientific claims and a panel of credible web-based news sources (eg, CNN[c] and BBC[d]) for news-related claims	English	4.5% inaccurate	H1N1	Exploratory
Sicilia et al [9]	800	Twitter	Annotator's agreement; relabeling in cases of disagreement; here, the definition for misinformation was "news items without a source"	English	Unknown	Zika	ML
Kalyanam et al [17]	47 million	Twitter	Type of hashtags	English	25% of the analyzed tweets were speculative	Ebola	Quantitative
Al-Rakhami and Al-Amri [18]	409,484	Twitter; keywords	Although they used coders, their definition of a rumor included lack of a source; hence, unconfirmed information was automatically classified as uncredible; in addition, tweets were classified by only 1 coder who checked interagreement on 20 tweets	Not noted, but the keywords were in English	70% uncredible	COVID-19	ML
Elhadad et al [19]	7486	Various websites	Fact-checking websites and official websites	English	21%	COVID-19	ML
Seltzer et al [20]	500	Instagram	Coders' agreement	English	23%	Zika	Exploratory
Ghenai et al [21]	26,728	Twitter	Defined keywords to the extracted tweets based on rumors identified from the WHO[e] website; then, the coders labeled the tweets	English	32%	Zika	ML

[a]ML: machine learning.

[b]CDC: Centers for Disease Control and Prevention.

[c]CNN: Cable News Network.

[d]BBC: British Broadcasting Corporation.

[e]WHO: World Health Organization.

Previous studies have focused on specific health issues and sometimes on specific types of rumors [8,18,19,21,22]. This suggests the need for a more general framework that can detect the accuracy of health information across known and previously unknown health conditions, such as during the outbreak of a previously unknown infectious disease.

Given the prevalent use of Twitter for the spreading of health information in Saudi Arabia [2,4-7,23,24], we aimed to inform the development of a new and more generic framework that is not bound to a specific disease or rumor type and detect the accuracy of a broad base of health-related tweets in Arabic.

Related Work

In this section, we review the methods used to label health-related tweets as either accurate or inaccurate to create labeled data sets. We also review previously proposed machine learning (ML) models for detecting the accuracy of health-related tweets, including deep learning (DL).

Methods Used to Label Health-Related Tweets

Studies addressing the accuracy of health-related tweets can be classified into 3 groups. The first group comprised studies that labeled health-related tweets according to the information they contained, regardless of the source of the information. The second group comprised studies that relied on external (fact-checking or very reputable) websites. The last group comprised studies that relied on various characteristics of the tweets or only on the source of the information to judge the accuracy of the tweets.

Regarding the concepts of accuracy and misinformation, Chou et al [25] defined *misinformation* as information that lacks scientific evidence. A more precise definition can be found in the study by Tan et al [26], where the authors defined inaccurate information or misinformation as “explicitly false,” according to what would be deemed incorrect by expert consensus. In the study by Nyhan and Reifler [27], the authors combined these definitions to describe misinformation or inaccurate health information as information that is not supported by clear evidence and expert opinion.

Studies relying on the opinions of experts seemed to indirectly or directly use these definitions to assess accuracy; however, it should be noted that, although misinformation is inaccurate, it is not necessarily intended to be so. In contrast, disinformation is information that is intentionally deceptive [28]. Examples of *opinions of experts* studies are included in Table 1 [8,10,11,14]. These involved labeling health-related tweets based on the opinions of health experts. The tweets were labeled as inaccurate or accurate by at least two experts. A third expert was typically involved when there was a disagreement between the original 2 experts: this expert cast the deciding vote for controversial tweets.

Vraga and Bode [29] criticized the abovementioned definition of misinformation, raising the point that there are many issues on which experts do not agree. However, they state that as long as there is more evidence supporting the information, the agreement rate between experts will increase. Taking a stricter approach, Albalawi et al [12] and Sharma et al [13] excluded tweets on which experts disagreed in an attempt to exclude uncertainty from their data sets. Table 1 summarizes these studies.

Unsurprisingly, studies that relied on expert opinion used relatively small data sets (ranging from 109 to 625 tweets) compared with studies that used other labeling methods (Table 1). Even those that used nonexperts but used manual coding (performed by nonexpert annotators) tended to work on a small sample of the data set [9,20].

The second group comprised studies that relied on an external website, such as a fact-checking website, to label the tweets. One such example is the study by Elhaddad et al [19], which relied on a fact-checking website to identify misleading information. A similar method was used by Ghenai et al [21], who relied on the website of the World Health Organization (WHO) to identify 6 rumors. From these rumors, they derived keywords to extract relevant tweets. The drawback of this method is that only tweets relevant to specific rumors were extracted; thus, the model was trained only on this limited number of rumors. Furthermore, these methods are highly language restricted: both studies referred to in Table 1 were performed in English, as mandated by the WHO website and the fact-checking website.

Other methods relied on various characteristics of the tweets or only on the source of the information without judging the actual information. For example, in the study by Kalyanam et al [17], the authors identified tweets as credible if they included hashtags that indicated that they originated from noted agencies or other reliable sources, and tweets were identified as speculative if they included hashtags that implied an increase in fear, rumors, or scams.

Similarly, Sicilia et al [9], Al-Rakhami and Al-Amri [18], and Chew and Eysenbach [16] defined credible tweets as tweets that have information from a confirmed, reliable source, such as the WHO, Centers for Disease Control, or another official health agency. This method differs from the method used by the second group mentioned previously as it first identified a tweet and then examined its source. In contrast, the methods in the second group first identified a trustworthy website and then used the information on the website to identify tweets of interest.

More generally, Yin et al [30] stated that a website is *trustworthy* if it provides correct information and suggests that information is likely to be true if it is provided by a trustworthy website. Studies that relied on trustworthy websites to identify rumors [9,18,21] seemed to follow this definition, even if they did not explicitly state it.

It should be noted that based on the data in Table 1, all Arabic studies that relied only on expert opinion [8,12,14] were small scale and qualitative; therefore, it would be impossible to scale them up. Notably, the percentage of inaccurate tweets for English studies that rely on expert opinions is in the range of 10% to 25%, whereas the corresponding range for Arabic studies is 31% to 50%. This finding suggests a greater occurrence of inaccurate health-related tweets in Arabic than in English.

ML Approaches

Of the 14 studies reported in Table 1, which analyzed the accuracy of health-related tweets in general, 6 (43%) proceeded to train an ML model to detect the accuracy of health information, as shown in Table 2.

Table 2. Summary of studies that developed MLa models to detect the accuracy of health-related information.

Study	ML approach	Results	Labeling type
Elhadad et al [19]	Deep learning multimodel, GRU[b], LSTM[c], and CNN[d]	99.99% (F_1 score)	Ground truth data from websites
Ghenai et al [21]	Random forest	94.5% (weighted average for F_1 score)	Crowdsource agreement but keywords are based on 4 WHO[e] website-identified rumors
Al-Rakhami and Al-Amri [18]	Ensemble learning and random forest+SVM[f]	97.8% (accuracy)	Single annotator only after confirming source
Zhao et al [10]	Random forest	84.4% (F_1 score)	Annotator vote; in addition, consulted an expert to validate misleading information
Sicilia et al [9]	Random forest	69.9% (F_1 score)	Agreement of a health expert
Saeed et al [8]	Random forest	83.5% (accuracy)	Agreement of a health expert

[a]ML: machine learning.

[b]GRU: gated recurrent unit.

[c]LSTM: long short-term memory.

[d]CNN: convolutional neural network.

[e]WHO: World Health Organization.

[f]SVM: support vector machine.

Studies reporting on training ML models included Elhadad et al [19] and Al-Rakhami and Al-Amri [18], who used ensemble learning on an English data set. Elhadad et al [19] used ensemble learning that involved multiple DL architectures, and Al-Rakhami and Al-Amri [18] trained ensemble models comprising traditional ML algorithms, such as support vector machine (SVM) and random forest (RF). Another similarity between these studies is the method used to identify misleading information. Elhadad et al [19] built their data set by extracting ground truth data and rumors from fact-checking websites. Al-Rakhami and Al-Amri [18] considered tweets credible if they have a reliable source and misleading otherwise. Both models reported a high level of accuracy (>97%), as shown in Table 2.

From Tables 1 and 2, it is clear that studies that relied on a fact-checking website [19,21] and studies that determined the accuracy of a tweet based on its source [18] obtained a high level of accuracy, possibly as these models were trained on relatively large data sets.

For example, Al-Rakhami and Al-Amri [18] trained their model using 409,484 tweets. However, automated labeling left open the possibility of incorrect labeling, and all these studies were conducted in English.

Most of the studies that developed ML models focused on outbreaks (4/6, 67% of studies). Studies that developed ML models for nonoutbreak conditions [8,10] obtained less accurate results compared with outbreak conditions. This might be because these nonoutbreak condition models were trained on a limited number of documents compared with the outbreak models. We also found that the level of accuracy obtained for nonoutbreak data sets was approximately 84% (Table 2). It is also notable that all of these studies trained an RF model.

Table 2 (and our associated literature review) suggests that recent advancements in DL have not been sufficiently applied to the detection of misleading Arabic health information. In our previous work, we have shown that DL architectures using word embedding as an input layer outperform other traditional ML models, such as SVM and naive Bayes, in the detection of Arabic health-related information on SM [31]; however, in this paper, we move past that to the classification of Arabic health-related tweets based on their accuracy.

Word embedding is a learned representation of words in natural language processing (NLP) [32]. Words with similar meanings typically have similar numbers in their vectors. The closer the words are in meaning, the shorter the distance between the 2 vectors representing them. One of the main criticisms of the word embedding approach is that it is considered context free; that is, the embedding of a word is not affected by its position in the sentence [33]. Hence, it is also referred to as static word embedding. However, in practice, the meaning of a word may depend on its position in a sentence.

In recent years, pretrained language models have been proven to work well for many NLP tasks, including entity recognition, language translation, and text classification [34]. Unlike static word embedding techniques, such as Skip-Gram and Continuous Bag of Words, language models can learn the context of the words and thus assign different values for the words depending on their context [33]. There are different types of language models, including contextual word vectors and embeddings from language models [33]. One of the most popular language models is the bidirectional encoder representations from transformers (BERT), which has been proven to perform well in text classification tasks.

The superiority of transformer models compared with other text classification methods is well documented, especially in the recent literature. Multiple studies have compared transformer models with other DL models [35-39], and the results showed that transformers outperformed the ML models, including different DL architectures and traditional ML models, such as SVMs and RF. This indicates the potential capability of transformers to better detect the accuracy of Arabic health information on SM.

Therefore, in this study, we aimed to contribute to this field by developing a data set of certified accurate or inaccurate Arabic health-related tweets and investigating the ability of the BERT or pretrained word embedding model to detect the accuracy of Arabic health-related tweets across a wide range of health-related issues.

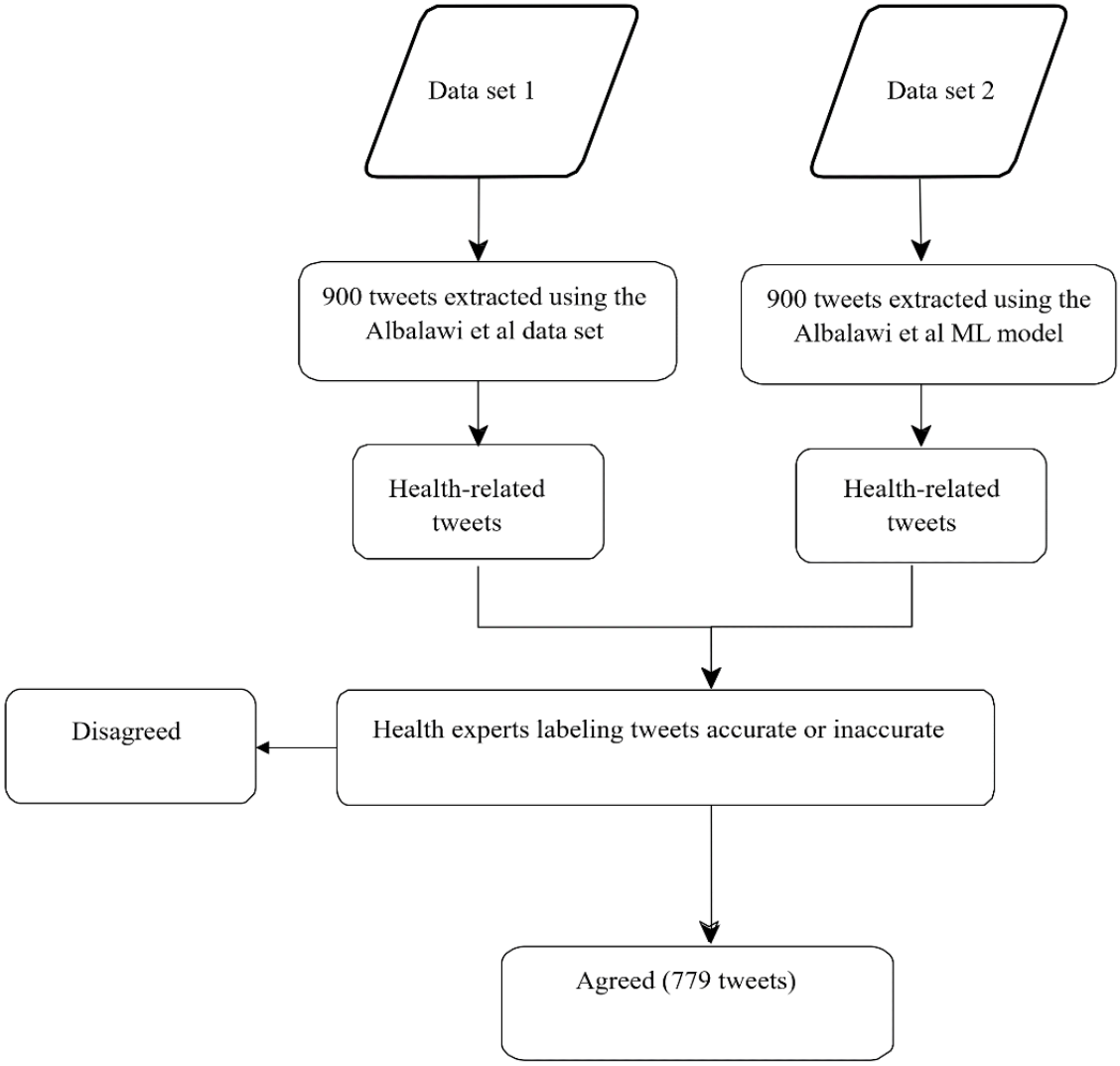
Methods

Overview

The empirical method comprised 2 parts. The first part addressed the extraction of health-related tweets using the model proposed in our previous study [31]. In that study, we used a health lexicon that focused more on general health keywords rather than specific outbreaks, as a recent study suggested that general health misinformation is more likely to spread than, for example, COVID-19 [40]. In contrast, Table 1 illustrates that most studies in this area focused on a specific domain or disease outbreak.

The extracted health-related tweets were labeled by health experts as either accurate or inaccurate. Figure 1 presents an overview of this portion of the study.

Figure 1. Overview of the process followed in labeling tweets as either accurate or inaccurate [31]. ML: machine learning.



In the second part, we propose 2 types of trustworthiness—detecting models to automatically classify health-related tweets as either accurate or inaccurate—and evaluate them: bidirectional long short-term memory (BLSTM) DL models and pretrained transformer language models.

Building Data Sets of Trustworthy Health-Related Tweets

In this study, we used 2 data sets containing health-related tweets. The first data set was the result of our previous study [31].

The first data set was extracted from 297,928 tweets posted between July 15 and August 31, 2019. Of these 297,928 tweets, 5000 (1.68%) were randomly sampled and labeled by 2 annotators as either *health-related* or *not health-related*. A third annotator resolved disagreements between the 2 annotators.

The first data set was extracted during the summer holidays in Saudi Arabia for 45 consecutive days. To assess generality, we extracted the second data set for a different timeframe: during *Hajj* and *Eid al Adha* (Muslim holy days) and during school days between March 1 and April 15, 2019. The second set of 900 tweets used the ML methodology proposed in the same study [31], as the availability of health professionals was constrained by the ongoing COVID-19 pandemic and the ML model derived in that study achieved a high-quality result (93% accuracy).

The methodology proposed in the study by Albalawi et al [31] comprised extracting tweets from a set of collected tweets with the help of a health lexicon and then further filtering out tweets not related to health with the help of an ML model. On the basis of the health lexicon, 217,702 tweets were extracted. Of the 217,702 tweets, we sampled 5000 (2.3%) tweets and applied the ML model to extract 900 (0.41%) health-related tweets.

Finally, we added 900 tweets from the second data set to 900 tweets sampled from the first data set and had those 1800 tweets labeled as either accurate or inaccurate health information by 2 medical physicians.

Labeling Accurate or Inaccurate Tweets

The physicians were asked to manually label each of the 1800 health-related tweets into one of the following categories: *accurate health information*, *inaccurate health information*, and *not sure about the accuracy*.

We followed the protocol of relying on the opinions of experts to define the accuracy of the information collected. Taking into account the points made by Vraga and Bode [29], every tweet was assessed by 2 experts, and a tweet was included in the final data set for this study only if both experts agreed on its accuracy; that is, we reduced uncertainty by excluding information that was not sanctioned by all experts (indeed, later show that between-physician reliability in this coding was limited, buttressing the need for increased certainty when using human classification, as stated by Vraga and Bode [32]). The *not sure* option was offered to the physicians to avoid forcing them to evaluate the tweets if they did not have enough relevant health knowledge to accurately evaluate them or if the tweets were ambiguous.

Although other studies invited a third annotator to resolve disagreements, our approach was stricter in reducing uncertainty in the data set by excluding tweets for which there was a disagreement between the 2 annotators. Of 1800 tweets, the 2 physicians agreed on 779 (43.3%) tweets, which were labeled as containing either accurate or inaccurate health information. The physicians disagreed on 9.1% (163/1800) of tweets. The remaining 47.7% (858/1800) of tweets were labeled as *unsure* by at least one physician. We dropped the tweets on which at least one of the physicians was unsure and used the remaining 779 tweets in our experiments.

Although the 779 tweets constituted a relatively small data set, most of the data sets constructed in the literature based on agreements between health experts were relatively small. As shown in Table 1, the highest number of health-related tweets judged by health experts in other studies was 625 in the study by Alnemer et al [14].

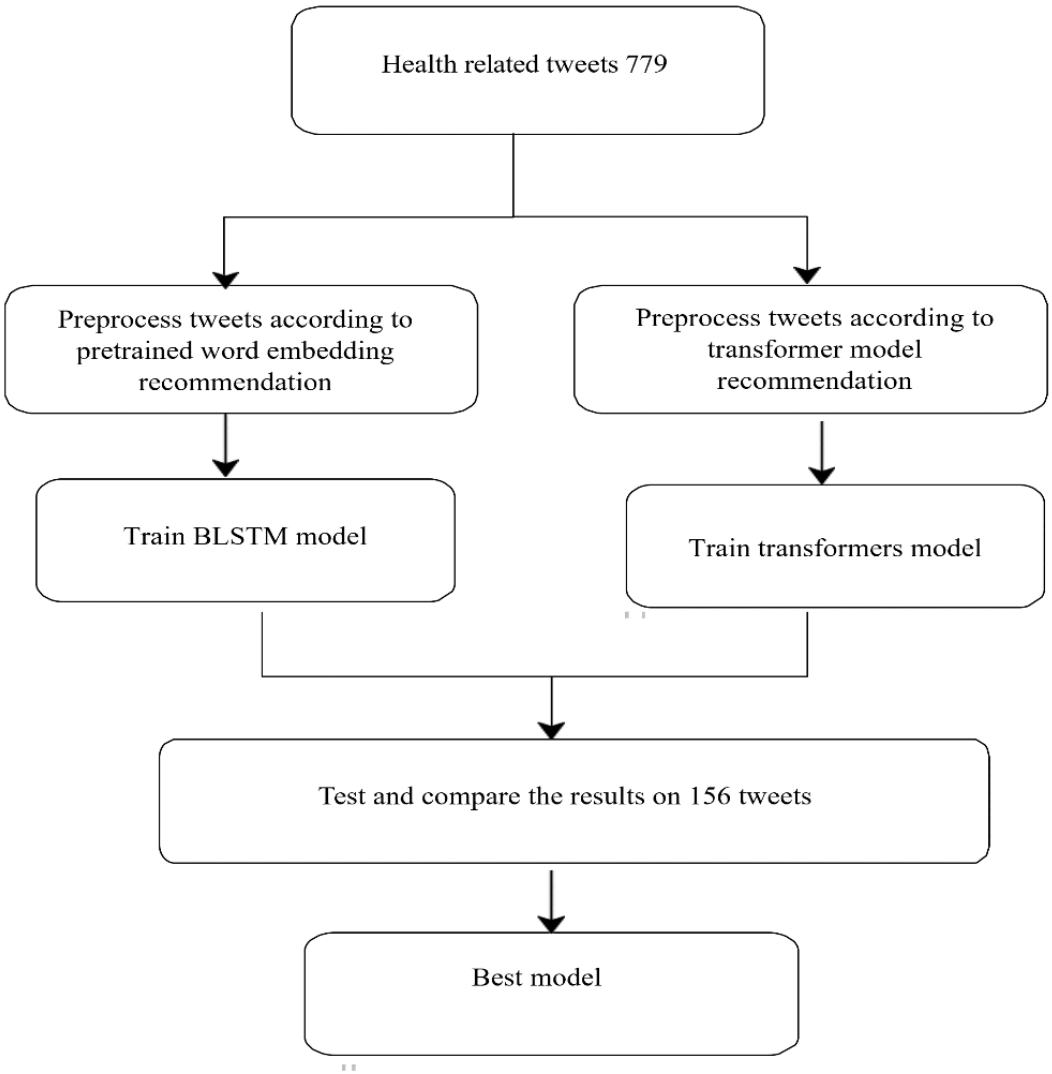
These 779 tweets, labeled as either accurate or inaccurate, can be found in Multimedia Appendix 1. Please note that we only share tweet IDs and labels as the Twitter policy prevents the content of the tweets from being redistributed. These tweet IDs can be used to obtain the text of tweets using the Twitter application programming interface [41].

Considered DL Models

Overview

After completing the annotation of the health-related tweets as either accurate or inaccurate, we trained 16 classification models, 7 (44%) of which used a BLSTM architecture with pretrained word embeddings as their input layers, and 9 (56%) of which used a pretrained transformer language model. Figure 2 illustrates the steps implemented during this stage. Further details are provided in the following sections.

Figure 2. Overview of the process used to train and select machine learning models. BLSTM: bidirectional long short-term memory.



The BLSTM Architecture

For 44% (7/16) of the trained models, we used a BLSTM architecture with pretrained word embeddings as the input layer. Long short-term memory (LSTM) is a type of recurrent neural network that takes advantage of dependencies between parts of the input sequence and can learn these dependencies. LSTM also preserves the information of past input. The BLSTM variation differs from LSTM because of its ability to learn the dependencies between past and future elements [42]. BLSTM has been found to perform well in many NLP tasks, including text classification [43]. The BLSTM model begins with input and embedding layers to which a dropout layer is added, followed by a BLSTM layer with another added dropout layer [31]. BLSTM has been shown to perform better than traditional ML models (SVM, naive Bayes, k-nearest neighbors, and logistic regression) and conventional neural networks in a previous study on detecting Arabic health-related tweets [31]. For the input layer, we used 7 pretrained word embedding models for Arabic [44-47]. It should be noted that AraVec, Mazajak, and ArWordVec come in 2 variations: Continuous Bag of Words and Skip-Gram and, finally, BLSTM fastText.

Transformer Models

BERT is a transformer language model that has shown superiority in many NLP tasks.

Different Arabic pretrained language models exist, which are based on transformers that have been developed recently by the Arabic NLP community. Most of these pretrained language models were built on top of the BERT-base model. Some of them also provided a version based on BERT-large.

The difference between BERT-base and BERT-large is that BERT-base uses 12 layers, 768 hidden layers, 12 heads, and approximately 136 million parameters, whereas the BERT-large model uses 24 layers, 1024 hidden layers, 16 heads, and approximately 370 million parameters [48]. All models may not leverage BERT-large as it is more difficult to train and comes with a higher computational cost than BERT-base [49].

Examples of pretrained Arabic language representation models that offer both base and large variants are ArabicBERT [50] and Transformer-based Model for Arabic Language Understanding (AraBERT) [51]. AraBERT was considered the first Arabic-specific transformer language model introduced in 2020 by Antoun et al [51]. In 2021, an updated version of

AraBERT was released [52]. AraBERT is considered one of the best transformer language models for NLP, outperforming other models for Arabic sentiment analysis [53]. AraBERT version 2 (AraBERTv2) preprocesses text using Farasa segmentation. Farasa segmentation involves breaking the words based on the prefix and suffix [54], whereas AraBERT version 0.2 (AraBERTv0.2) preprocesses the text without using Farasa segmentation. In this study, we experimented with these 6 models: AraBERTv2, AraBERTv0.2, and ArabicBERT in both variants of BERT (base and large).

In addition to 6 models, we also investigated 3 other state-of-the-art pretrained language models, namely QARiB [55], MARBERT, and ARBERT [56], which are based only on BERT-base. These models reportedly perform well on text classification tasks [50,55-57]. Table 3 summarizes the characteristics of the pretrained language models used in this study.

Table 3. Pretrained language models.

Name	Basis	Size	Corpus
ARBERT [56]	BERT[a]-base	61 GB of MSA[b] text (6.5 billion tokens)	• Books and news (news and Wikipedia articles)
MARBERT [56]	BERT-base	128 GB of text (15.6 billion tokens)	• 1 billion Arabic tweets
QARiB [55]	BERT-base	14 billion tokens; vocabulary: 64,000	• 420 million tweets and approximately 180 million sentences of text from Arabic Giga Word, Abulkhair Arabic Corpus, and OPUSc
ArabicBERT [50]	BERT-base and BERT-large	95 GB of text and 8.2 billion words	• Arabic OSCARd version, Wikipedia, and other resources
AraBERTv0.2[e] [52]	BERT-base and BERT-large	77 GB, 200,095,961 lines, 8,655,948,860 words, or 82,232,988,358 characters	• OSCAR unshuffled and filtered • Arabic Wikipedia articles • The 1.5 billion words Arabic Corpus • The OSIANf corpus • Assafir news articles
AraBERTv2[g] [52]	BERT-base and BERT-large	77 GB, 200,095,961 lines, 8,655,948,860 words, or 82,232,988,358 characters	• OSCAR, unshuffled and filtered • Arabic Wikipedia articles • The 1.5 billion words Arabic corpus • The OSIAN corpus • Assafir news articles

[a]BERT: bidirectional encoder representations from transformers.

[b]MSA: Modern Standard Arabic.

[c]OPUS: open parallel corpus.

[d]OSCAR: Open Superlarge Crawled Aggregated corpus.

[e]AraBERTv0.2: Transformer-based Model for Arabic Language Understanding version 0.2version 0.2.

[f]OSIAN: Open Source International Arabic News.

[g]AraBERTv2: Transformer-based Model for Arabic Language Understanding version 0.2 version 2.

Evaluation Metrics

The F_1 score, recall, precision, accuracy, and macroaverage of the F_1 score were used to evaluate the ML models, as detailed in Textbox 1. The macroaveraged F_1 score is the averaged F_1 score across all classes, which are accurate and inaccurate health-related tweets [58].

Textbox 1. Metrics used to evaluate the machine learning models.

Recall

- True positives / (true positives + false negatives)

Precision

- True positives / (true positives + false positives)

F_1 score

- (2 × precision × recall) / (precision + recall)

Accuracy

- (true positives + true negatives) / total sample

Macroaveraged F_1 score

-  **(1)**, where N is the number of classes

Preprocessing Data

In this study, text was preprocessed following the procedure outlined by the authors of the corresponding pretrained word embedding models. Li et al [59] found that this is the best text preprocessing practice when working with pretrained word embeddings. Similarly, for all pretrained word embedding models [44-47] and pretrained language models [50,52,55,56], we followed the steps provided by the original studies.

Of the 779 tweets, we split the data set into training, validation, and test data sets in ratios of 507 tweets (65.1%) for training, 116 tweets (14.9%) for validating the model, and 156 tweets (20%) for testing.

Ethics Approval

This study did not require institutional review board approval from the Science and Engineering Research committee at the University of Limerick because ethical approval is not required for publicly available data. It should be emphasized, during the study, that any associated text that can be used to identify the authors of the tweets has been removed from the text (eg, @name, user ID).

Results

Data Set Description

The κ coefficient for all categories was 0.377, which is in fair agreement according to Cohen [60]. However, the benchmark scale proposed by Fleiss et al [61] to evaluate the agreement indicates that such a coefficient is poor (<0.40=poor, 0.40-0.75=intermediate to good, and >0.75=excellent). Given the low κ coefficients across the 3 categories, we considered only cases where both physicians were explicitly in agreement, as they were on 779 tweets from the original data sets.

Of the 1021 tweets that were excluded, 874 (48.6%) were labeled *not sure* by at least one physician, and in the case of 147 (14.4%) tweets, the physicians disagreed regarding the accuracy of the tweets.

Of the 779 tweets physicians agreed on in our data set, 296 (38%) were labeled as inaccurate and 483 (62%) were labeled as accurate. This finding is similar to the inaccuracies reported in other studies (Table 1).

Textbox 2 presents examples of accurate and inaccurate health-related tweets. As can be seen from the tweets in the textbox, they cover a wide range of topics, including but not limited to psychology and cancer. Interestingly, in the third accurate tweet example, the difficulty for nonexperts in discerning accurate from inaccurate health information is illustrated, as advice against taking antidiarrhea drugs in the event of food poisoning is slightly counterintuitive.

Textbox 2. Examples of inaccurate and accurate health-related tweets.

Accurate

- "Tomorrow enjoys the feast. and get closer to God with your sacrifice

  And eat but do not extravagant and feed the contented and be merciful as God has commanded you

  Eating too much red meat might:

  Raise the level of triglycerides

  Raise cholesterol

  Increase uric salt in the blood Increases gout attacks in the joints"

- "Symptoms of social phobia

  Sometimes, social phobia can be accompanied by physical signs and symptoms, which may include:

  Flashness

  Rapid heart palpitations

  Shivering and sweating

  Upset stomach or nausea

  Difficulty catching breath

  Dizziness or lightheadedness

  Feeling like your mind has gone blank

  Muscle tension"

- "In the event of food poisoning, please take care not to use antidiarrheal medicines, as they may worsen the condition"
- "Hemoglobin is a group of proteins in red blood cells whose function is to transport oxygen from the lungs to the body, return carbon dioxide from the body, and transport it to the lungs and get rid of it through breathing.

Iron is an important element and enters the composition of hemoglobin, so if iron deficiency, hemoglobin decreases, and anemia occurs."

- "Among the ways to prevent lung cancer:

  Stay away from smoking

  Avoid passive smoking

  Avoid carcinogenic and radioactive materials."

Inaccurate

- "Scientific research,

  The research says that Zamzam water bears the name (water), but it differs radically from water compounds, as all the waters of the world belong to the acidic compound, except for (Zamzam water).

  It is (alkaline!) Glory be to God. There is no other alkaline water on the face of the earth. So, when you drink it in abundance, the human body has a strong immunity against viruses!!"

- "When Western scholars searched for the causes of mental illness, they found only two reasons (fear and sadness) fear of the future and sadness of the past, both of which are the opposite of happiness."
- "Did you know that a 5-minute tantrum is so stressful that it weakens the immune system for more than 6 hours"
- "Cupping helps smokers to quit smoking or reduce the negative impact on the body through:

  Removing excess hemoglobin from the body by excreting aging red blood cells, and thus the disappearance of the pathological symptoms of high hemoglobin caused by smoking"

- "Just a spoonful of cinnamon daily:

  Rich in anti-inflammatory and antioxidants

  Prevents all types of cancer

  Prevents heart disease

  Anti-diabetes"

Some tweets claimed the benefits of some traditional foods and spices. For example, some tweets promoted *Zamzam* (holy water for Muslims), claiming there was scientific research that stated that it could strengthen the human immune system; experts classified the information as inaccurate.

In addition, the examples of accurate tweets presented here suggest that accurate health-related tweets tend to be more preventive in nature, a finding supported by the wider sampling of accurate tweets. As shown in Textbox 2, the accurate tweets advised users to stop eating too much red meat as it causes gout or increases cholesterol, stop smoking to prevent lung cancer, and stop taking anti-inflammatory drugs in the event of food poisoning. In contrast, as noted earlier, inaccurate tweets promoted natural and alternative medicine such as curbing eating and drinking *Zamzam* water for their health benefits. An interesting example was in relation to cancer, where accurate tweets advised readers to stop smoking; however, some of the inaccurate tweets were also preventive, and they advised taking a spoonful of cinnamon to prevent all types of cancer.

DL Models

In terms of the comparison of models, we observed that overall, BERT models performed better than BLSTM models based on the accuracy and the F_1 score for both classes (when referring to the metric accuracy in this section, we will call it *model accuracy* to disambiguate it from the accurate or inaccurate classification). Overall, AraBERTv0.2-large performed better than all other models. Specifically, the best model was AraBERTv0.2-large (macro F_1 score 87%), followed by AraBERTv2-large (macro F_1 score 86%) and AraBERTv0.2-base (macro F_1 score 85%), as shown in Table 4. These findings hide larger but still small variations in the precision and recall scores of individual techniques for inaccurate and accurate tweets. For example, although AraBERTv0.2-base achieved a recall of 78% for inaccurate tweets, AraBERTv0.2-large achieved a recall of >83%.

The results also suggest that, in general, BERT-large models tended to be better at detecting inaccurate tweets than the BERT-base models. The large AraBERTv2, AraBERTv0.2, and ArabicBERT models performed better than their base versions at detecting inaccurate health tweets, as shown in Table 4. In contrast, the BERT-base models might be better at detecting accurate tweets, except for the AraBERTv2, whose large and base versions performed similarly.

Of the pretrained word embeddings, the results in Table 4 show that Mazajak Skip-Gram is the best based on *model accuracy* and F_1 score.

Table 4. Comparison of the performance of machine learning models for detecting the accuracy of health-related tweets.

Model and class	Precision	Recall	F_1 score	Macroaverage	Model accuracy
AraBERTv2[a]-base					
Inaccurate	0.804	0.7627	0.7826	0.8279	0.8397
Accurate	0.86	0.8866	0.8731	0.8279	0.8397
AraBERTv2-large					
Inaccurate	0.8276	0.8136[b]	0.8205[b]	0.8564[b]	0.8654[b]
Accurate	0.8878	0.8969	0.8923	0.8564[b]	0.8654[b]
AraBERTv0.2[c]-base					
Inaccurate	0.8519	0.7797	0.8142	0.8543[b]	0.8654[b]
Accurate	0.8725	0.9175	0.8945	0.8543[b]	0.8654[b]
AraBERTv0.2-large					
Inaccurate	0.8448	0.8305[d]	0.8376[d]	0.8701[d]	0.8782[d]
Accurate	0.898[d]	0.9072	0.9025[d]	0.8701[d]	0.8782[d]
MARBERT					
Inaccurate	0.7759	0.7627	0.7692	0.8154	0.8269
Accurate	0.8571	0.866	0.8615	0.8154	0.8269
ARBERT					
Inaccurate	0.7903	0.8305[d]	0.8099	0.8447	0.8526
Accurate	0.8936	0.866	0.8796	0.8447	0.8526
QARiB					
Inaccurate	0.7797	0.7797	0.7797	0.8228	0.8333
Accurate	0.866	0.866	0.866	0.8228	0.8333
ArabicBERT[e]-large					
Inaccurate	0.8654	0.7627	0.8108	0.8532	0.8654[b]
Accurate	0.8654	0.9278[b]	0.8955[b]	0.8532	0.8654[b]
ArabicBERT-base					
Inaccurate	0.8913[d]	0.6949	0.781	0.83492	0.8525
Accurate	0.8364	0.9485[d]	0.8889	0.83492	0.8525
BLSTM[f] Mazajak CBOW[g]					
Inaccurate	0.7719	0.7458	0.7586	0.8079	0.8205
Accurate	0.8485	0.866	0.8571	0.8079	0.8205
BLSTM Mazajak Skip-Gram					
Inaccurate	0.8542	0.6949	0.7664	0.8222	0.8397
Accurate	0.8333	0.9278[b]	0.8780	0.8222	0.8397
BLSTM ArWordVec Skip-Gram					
Inaccurate	0.8261	0.6441	0.7238	0.7919	0.8141
Accurate	0.8091	0.9175	0.8148	0.7919	0.8141
BLSTM ArWordVec CBOW					
Inaccurate	0.7925	0.7119	0.75	0.805	0.8205
Accurate	0.835	0.8866	0.86	0.805	0.8205

Model and class	Precision	Recall	F_1 score	Macroaverage	Model accuracy
BLSTM AraVec CBOW					
Inaccurate	0.6865	0.7797	0.7302	0.7737	0.7821
Accurate	0.8571	0.866	0.8172	0.7737	0.7821
BLSTM AraVec Skip-Gram					
Inaccurate	0.7313	0.8305[d]	0.7777	0.8136	0.8205
Accurate	0.8144	0.8144	0.8494	0.8136	0.8205
BLSTM fastText					
Inaccurate	0.8158	0.5254	0.6392	0.7382	0.7756
Accurate	0.7627	0.9278[b]	0.8372	0.7382	0.7756

[a]AraBERTv2: Transformer-based Model for Arabic Language Understanding version 2.

[b]Represents the second-best value.

[c]AraBERTv0.2: Transformer-based Model for Arabic Language Understanding version 0.2.

[d]Indicates the best value.

[e]BERT: bidirectional encoder representations from transformers.

[f]BLSTM: bidirectional long short-term memory.

[g]CBOW: Continuous Bag of Words.

Discussion

Principal Findings

As noted earlier, the examples given in the *Results* section showed that accurate tweets were more focused on preventive medicine, whereas inaccurate tweets were more focused on alternative and natural medicine. However, it could be argued that this is because of the keywords used in extracting and filtering the tweets or because of the selected tweet examples. Nevertheless, a previous study mentioned that the prevalence of natural alternatives and alternative medicine compared with medicine provided by the health care system [62] may be harmful. To illustrate the importance of this with respect to specific patients, there was a reported case of a patient with cancer who took alternative medicine promoted on SM, which caused the hospital to temporarily stop her cancer treatment to repair the damage caused by that medicine [63]. At a more general level, going forward, insights such as these could provide additional levers with which to detect inaccurate health tweets.

The results of BLSTM with pretrained word embedding models (AraVec, Skip-Gram, and Mazajak) are comparable with the results of some BERT models, including MARBERT, QARiB, and ArabicBERT-large. Indeed, this has been previously reported in the literature, where MARBERT and QARiB outperformed some of the other transformer models, such as ArabicBERT and AraBERT [55,56]. Again, a takeaway from this is that pretrained word embeddings might outperform pretrained BERT models in this first comparative study directed at Arabic. There is no guaranteed best model between pretrained word embeddings and pretrained transformer models for this language.

However, in general, the results showed the superiority of the BERT models over BLSTM with pretrained word embedding models. Overall, 19 best or second-best results were obtained by the 9 BERT-based approaches, whereas only 3 best or second-best results were obtained by the 7 pretrained word embedding models.

Most models performed better at detecting accurate health tweets than inaccurate tweets. The detection rate (recall) for accurate tweets ranged from 0.9485 to 0.8144. This means that most of the models missed only approximately 5% to 19% of the accurate tweets, which is a promising result. In contrast, the detection rate for inaccurate tweets was lower and had a wider range, from 0.8305 to 0.5254, implying that the best models missed up to 17% of inaccurate tweets. This is concerning as we would like to successfully identify all inaccurate tweets, and even the best model missed 17% of them.

The flip side of this is precision: how many accurate or inaccurate tweets identified by the technique are actually accurate or inaccurate. In terms of inaccurate tweets, the approaches ranged from 0.7759 to 0.89130—quite a large span, which means that if the wrong technique is chosen, approximately one-quarter of the tweets identified as inaccurate is incorrectly classified. Probably, more of a concern is the number of tweets identified as accurate that are not. Similarly, here, the span ranged from 0.8913 to 0.7627, again implying that if the wrong technique is chosen, this could be problematic.

Some models that had high detection rates for accurate health tweets could have low detection rates for inaccurate tweets. For example, the ArabicBERT-base and BLSTM fastText models were the best and second best for accurately detecting tweets, with success rates of 0.9485 and 0.9278, respectively. However, in detecting inaccurate tweets, BLSTM fastText had the lowest detection rate (52%) and the ArabicBERT-Base model had the second-lowest detection rate (69%). In other words, a practitioner who uses the best model for identifying accurate health tweets might miss approximately 30% to 48% of inaccurate tweets.

Similarly, the ARBERT and AraVec Skip-Gram models performed similarly to the AraBERTv0.2-large model in terms of precision when detecting inaccurate health-related tweets; however, these 2 models did not perform as well on the other metrics. For example, the AraVec Skip-Gram model had the second-lowest rate of *model accuracy* in classifying accurate tweets as inaccurate. Although the ARBERT model performed well compared with the BLSTM models, with regard to classifying accurate tweets as inaccurate, it had the third-lowest rate of *model accuracy* among the 9 BERT models tested in this study. In other words, the ARBERT models incorrectly classified accurate tweets as inaccurate at a higher rate than the 6 other BERT models, as shown in Table 4.

Ideally, a technique would provide high precision in both identification and recall; however, this did occur in the data set for accurate or inaccurate tweets. AraBERTv0.2-large came closest in this regard with high-accuracy tweet precision and recall, best recall for inaccurate tweets, and suboptimal precision for inaccurate tweets. Similarly, AraBERTv2-large performed quite well across accurate tweets but did not perform quite well on inaccurate tweets.

However, these models (AraBERTv0.2-large and AraBERTv2-large) consume relatively more resources, being based on BERT-large. Among the base models, AraBERTv0.2-base has an F_1 score of 0.8543, which is good, and also has a similar *model accuracy* to AraBERTv2-large. These models can be considered as an alternative if resources are an important consideration.

Regarding the performance of pretrained word embeddings, we found that Mazajak Skip-Gram was the best. We made the same observation in our previous work on the detection of health-related tweets [31].

Finally, with respect to the accuracy of the best model in our study (ie, AraBERTv0.2-large), our results are satisfactory when compared with the results of previous studies [8-10] that make use of expert opinion. The F_1 score of our best model was 87%, whereas the best F_1 score reported in the study by Zhao et al [10] was 84%, as shown in Table 2. Furthermore, although these previous studies targeted a specific health topic (such as cancer [8] or autism [10]), we used a data set of tweets on a wide range of health care topics, suggesting that it would be more difficult to classify our data set.

It should be noted that all 3 studies with *model accuracy* or F_1 scores >90% did not rely on expert opinion (Tables 1 and 2). In addition, 2 of these 3 studies [18,19] targeted a specific outbreak condition (COVID-19), and their models were trained on a larger data set (eg, Al-Rakhami and Al-Amri [18] trained their model on 409,484 tweets). For the third study [21], the keywords used to extract initial tweets were derived from 6 preidentified rumors related to Zika. The size and nature of the data used to train these models might explain why they seemed to achieve better accuracy than the model proposed here. In this study, we trained a model to detect the accuracy of generic health-related information, making the approach applicable to tweets that are more or less categorical in their labeling (as illustrated in the samples in Textbox 2).

Limitations

This study only considered tweets agreed upon by experts. Although this helps us reduce the uncertainty in our data set, it might be a limitation as the model is not trained or tested on tweets that are more marginal—tweets about which the experts are unsure.

One of the strengths of this model is that it was trained on general health-related tweets. The accuracy of the model for each health condition or topic may vary, and future studies should evaluate the model for specific health topics.

All models used here are language dependent and might not be directly applicable to other languages. However, there are BERT alternatives for many languages, and there is evidence that BERT outperforms word embedding-based models. Therefore, we believe that this model could perform similarly in other languages.

Regarding the metrics used to evaluate the models, it should be noted that the F_1 measure has been subjected to some criticism. Although we showed the F_1 score for both classes (accurate and inaccurate health tweets), it should be noted that the measure gives equal importance to both classes (accurate and inaccurate health tweets). Moreover, the F_1 score generally does not consider true negatives in its equation [64,65].

Conclusions

The goal of this study was to develop and evaluate a state-of-the-art ML model for detecting the medical trustworthiness of health-related tweets in Arabic. To achieve this, we first constructed a labeled data set to train the classifiers. We then compared 2 different DL approaches for training a classification model, namely, 6 pretrained word embedding models as an input model for BLSTM and 11 pretrained transformer language models. The percentage of inaccurate health tweets in the data is approximately 38% (296/799), which is comparable with previous studies that used data sets with a number of inaccurate health-related tweets in the range of 30% to 50%. Our AraBERTv0.2-large model achieved 87.7% *model accuracy* on the test data set, which is satisfactory. Overall, our results clearly indicate that the AraBERTv0.2-large model outperforms the other models in detecting the medical accuracy of health-related tweets.

This study established an ML model to identify the accuracy of health-related tweets in response to the proliferation of health misinformation on SM. Although misinformation detection has been researched, only 1 study was concerned with detecting the accuracy of Arabic health-related tweets, and it was only for a specific topic (cancer). Furthermore, no DL model has been evaluated in prior studies to detect the accuracy of Arabic health-related tweets. In this study, we used a more extensive data set to develop a more general model using state-of-the-art ML models that have not been implemented before for this type of problem.

The potential of such work cannot be overstated. If a robust model can be built, it will allow for the detection and dissemination of accurate tweets. Similarly, this would allow for the flagging of inaccurate tweets. Both measures would

significantly improve health information dissemination on Twitter. However, it should be noted that although this work will improve the situation, it will still inaccurately classify 13% of the tweets.

Moreover, the examples in Textbox 2 imply disparities between accurate and inaccurate information in terms of the topics covered across the data set—a trend supported by the informal sampling of that data set. Accurate tweets seem to be more preventive, whereas inaccurate health tweets seem to promote *natural* and alternative medicine. Thus, it might be more feasible to develop a model for detecting health topics in combination with a model for detecting the accuracy of health information and thus improving accuracy.

To further improve the accuracy of the developed model, ensemble learning can yield better results by combining models that perform well (ArabicBERT-large, ARBERT, and AraVec Skip-Gram). However, ArabicBERT and AraBERTv0.2 were trained on a similar corpus, as shown in Table 3. Another approach could be to combine models pretrained on different corpora, such as ArabicBERT-large and MARABER (ArabicBERT pretrained on Wikipedia articles and news articles; MARBERT pretrained on 1 billion tweets).

Acknowledgments

The authors would like to thank all physicians who participated in the process of evaluating the health tweets in this study. This work was supported, in part, by Taibah University, Al-Ula, Saudi Arabia, and by a grant from the Science Foundation Ireland (grant 13/RC/2094).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Tweet IDs with their labels used in the study.
[XLSX File (Microsoft Excel File), 26 KB - formative_v6i6e34834_app1.xlsx]

References

1. Ott BL. The age of Twitter: Donald J. Trump and the politics of debasement. Critical Stud Media Commun 2016 Dec 23;34(1):59-68. [doi: 10.1080/15295036.2016.1266686]
2. El Tantawi M, Bakhurji E, Al-Ansari A, AlSubaie A, Al Subaie HA, AlAli A. Indicators of adolescents' preference to receive oral health information using social media. Acta Odontol Scand 2019 Apr;77(3):213-218. [doi: 10.1080/00016357.2018.1536803] [Medline: 30632864]
3. Tang Y, Hew KF. Using Twitter for education: beneficial or simply a waste of time? Comput Educ 2017 Mar;106:97-118. [doi: 10.1016/j.compedu.2016.12.004]
4. Justinia T, Alyami A, Al-Qahtani S, Bashanfar M, El-Khatib M, Yahya A, et al. Social media and the orthopaedic surgeon: a mixed methods study. Acta Inform Med 2019 Mar;27(1):23-28 [FREE Full text] [doi: 10.5455/aim.2019.27.23-28] [Medline: 31213739]
5. Hamasha AA, Alghofaili N, Obaid A, Alhamdan M, Alotaibi A, Aleissa M, et al. Social media utilization among dental practitioner in Riyadh, Saudi Arabia. Open Dentistry J 2019 Feb 28;13(1):101-106. [doi: 10.2174/1874210601913010101]
6. Jamal A, Khan SA, AlHumud A, Al-Duhyyim A, Alrashed M, Bin Shabr F, et al. Association of online health information-seeking behavior and self-care activities among type 2 diabetic patients in Saudi Arabia. J Med Internet Res 2015 Aug 12;17(8):e196 [FREE Full text] [doi: 10.2196/jmir.4312] [Medline: 26268425]
7. Omair MA, AlOhaly RY, Alashgar LM. Awareness and misconceptions of female students in King Saud University on systemic lupus erythematosus. Rheumatology (Sunnyvale) 2015;05(03):165. [doi: 10.4172/2161-1149.1000165]
8. Saeed F, Yafooz W, Al-Sarem M, Hezzam E. Detecting health-related rumors on Twitter using machine learning methods. Int J Advanced Comput Sci Application 2020;11(8). [doi: 10.14569/ijacsa.2020.0110842]
9. Sicilia R, Giudice S, Pei Y, Pechenizkiy M, Soda P. Health-related rumour detection on Twitter. In: Proceedings of the 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2017 Presented at: 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); Nov 13-16, 2017; Kansas City, MO, USA. [doi: 10.1109/bibm.2017.8217899]
10. Zhao Y, Da J, Yan J. Detecting health misinformation in online health communities: incorporating behavioral features into machine learning based approaches. Inf Process Manag 2021 Jan;58(1):102390. [doi: 10.1016/j.ipm.2020.102390]
11. Swetland SB, Rothrock AN, Andris H, Davis B, Nguyen L, Davis P, et al. Accuracy of health-related information regarding COVID-19 on Twitter during a global pandemic. World Med Health Policy 2021 Jul 29;13(3):503-517 [FREE Full text] [doi: 10.1002/wmh3.468] [Medline: 34540337]
12. Albalawi Y, Nikolov NS, Buckley J. Trustworthy health-related tweets on social media in Saudi Arabia: tweet metadata analysis. J Med Internet Res 2019 Oct 08;21(10):e14731 [FREE Full text] [doi: 10.2196/14731] [Medline: 31596242]

13. Sharma M, Yadav K, Yadav N, Ferdinand KC. Zika virus pandemic-analysis of Facebook as a social media health information platform. Am J Infect Control 2017 Mar 01;45(3):301-302. [doi: 10.1016/j.ajic.2016.08.022] [Medline: 27776823]
14. Alnemer KA, Alhuzaim WM, Alnemer AA, Alharbi BB, Bawazir AS, Barayyan OR, et al. Are health-related tweets evidence based? Review and analysis of health-related tweets on twitter. J Med Internet Res 2015 Oct 29;17(10):e246 [FREE Full text] [doi: 10.2196/jmir.4898] [Medline: 26515535]
15. Sell TK, Hosangadi D, Trotochaud M. Misinformation and the US Ebola communication crisis: analyzing the veracity and content of social media messages related to a fear-inducing infectious disease outbreak. BMC Public Health 2020 May 07;20(1):550 [FREE Full text] [doi: 10.1186/s12889-020-08697-3] [Medline: 32375715]
16. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PLoS One 2010 Nov 29;5(11):e14118 [FREE Full text] [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]
17. Kalyanam J, Velupillai S, Doan S, Conway M, Lanckriet G. Facts and fabrications about Ebola: a Twitter based study. In: Proceedings of the KDD '15. 2015 Presented at: KDD '15; Aug 10 – 13, 2015; Sydney, Australia.
18. Al-Rakhami MS, Al-Amri AM. Lies kill, facts save: detecting COVID-19 misinformation in Twitter. IEEE Access 2020;8:155961-155970. [doi: 10.1109/access.2020.3019600]
19. Elhadad M, Li K, Gebali F. An ensemble deep learning technique to detect COVID-19 misleading information. In: Advances in Networked-Based Information Systems. Cham: Springer; 2021.
20. Seltzer E, Horst-Martz E, Lu M, Merchant R. Public sentiment and discourse about Zika virus on Instagram. Public Health 2017 Sep;150:170-175. [doi: 10.1016/j.puhe.2017.07.015] [Medline: 28806618]
21. Ghenai A, Mejova Y. Catching zika fever: application of crowdsourcing and machine learning for tracking health misinformation on Twitter. In: Proceedings of the 2017 IEEE International Conference on Healthcare Informatics (ICHI). 2017 Presented at: 2017 IEEE International Conference on Healthcare Informatics (ICHI); Aug 23-26, 2017; Park City, UT, USA. [doi: 10.1109/ichi.2017.58]
22. Wang Y, McKee M, Torbica A, Stuckler D. Systematic literature review on the spread of health-related misinformation on social media. Soc Sci Med 2019 Nov;240:112552 [FREE Full text] [doi: 10.1016/j.socscimed.2019.112552] [Medline: 31561111]
23. Alhaddad MS. The use of social media among Saudi residents for medicines related information. Saudi Pharm J 2018 Dec;26(8):1106-1111 [FREE Full text] [doi: 10.1016/j.jsps.2018.05.021] [Medline: 30510470]
24. Alsobayel H. Use of social media for professional development by health care professionals: a cross-sectional web-based survey. JMIR Med Educ 2016 Sep 12;2(2):e15 [FREE Full text] [doi: 10.2196/mededu.6232] [Medline: 27731855]
25. Chou WS, Oh A, Klein WM. Addressing health-related misinformation on social media. JAMA 2018 Dec 18;320(23):2417-2418. [doi: 10.1001/jama.2018.16865] [Medline: 30428002]
26. Tan AS, Lee C, Chae J. Exposure to health (mis)information: lagged effects on young adults' health behaviors and potential pathways. J Commun 2015 Jul 06;65(4):674-698. [doi: 10.1111/jcom.12163]
27. Nyhan B, Reifler J. When corrections fail: the persistence of political misperceptions. Polit Behav 2010 Mar 30;32(2):303-330. [doi: 10.1007/s11109-010-9112-2]
28. Pool J, Fatehi F, Akhlaghpour S. Infodemic, misinformation and disinformation in pandemics: scientific landscape and the road ahead for public health informatics research. Stud Health Technol Inform 2021 May 27;281:764-768. [doi: 10.3233/SHTI210278] [Medline: 34042681]
29. Vraga EK, Bode L. Defining misinformation and understanding its bounded nature: using expertise and evidence for describing misinformation. Political Commun 2020 Feb 06;37(1):136-144. [doi: 10.1080/10584609.2020.1716500]
30. Yin X, Han J, Yu P. Truth discovery with multiple conflicting information providers on the web. IEEE Trans Knowl Data Eng 2008 Jun;20(6):796-808. [doi: 10.1109/tkde.2007.190745]
31. Albalawi Y, Buckley J, Nikolov NS. Investigating the impact of pre-processing techniques and pre-trained word embeddings in detecting Arabic health information on social media. J Big Data 2021 Jul 02;8(1):95 [FREE Full text] [doi: 10.1186/s40537-021-00488-w] [Medline: 34249602]
32. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. In: Proceedings of Workshop at ICLR. 2013 Presented at: Workshop at ICLR; May 2-4, 2013; Scottsdale, Arizona, USA URL: https://ui.adsabs.harvard.edu/abs/2013arXiv1301.3781M
33. Qiu X, Sun T, Xu Y, Shao Y, Dai N, Huang X. Pre-trained models for natural language processing: a survey. Sci China Technol Sci 2020 Sep 15;63(10):1872-1897. [doi: 10.1007/s11431-020-1647-3]
34. Azzouza N, Akli-Astouati K, Ibrahim R. TwitterBERT: framework for Twitter sentiment analysis based on pre-trained language model representations. In: Emerging Trends in Intelligent Computing and Informatics. Cham: Springer; 2020.
35. Abu Farha I, Magdy W. A comparative study of effective approaches for Arabic sentiment analysis. Inf Process Manag 2021 Mar;58(2):102438. [doi: 10.1016/j.ipm.2020.102438]
36. Al-Twairesh N. The evolution of language models applied to emotion analysis of Arabic Tweets. Information 2021 Feb 17;12(2):84. [doi: 10.3390/info12020084]
37. El-Alami F, Ouatik El Alaoui S, En Nahnahi N. Contextual semantic embeddings based on fine-tuned AraBERT model for Arabic text multi-class categorization. J King Saud University Comput Inf Sci 2021 Feb (forthcoming). [doi: 10.1016/j.jksuci.2021.02.005]

38. El-Razzaz M, Fakhr MW, Maghraby FA. Arabic gloss WSD using BERT. Applied Sci 2021 Mar 13;11(6):2567. [doi: 10.3390/app11062567]
39. Gomez-Perez J, Denaux R, Garcia-Silva A. Understanding word embeddings and language models. In: A Practical Guide to Hybrid Natural Language Processing. Cham: Springer; 2020.
40. Broniatowski DA, Kerchner D, Farooq F, Huang X, Jamison AM, Dredze M, et al. Twitter and Facebook posts about COVID-19 are less likely to spread misinformation compared to other health topics. PLoS One 2022 Jan 12;17(1):e0261768 [FREE Full text] [doi: 10.1371/journal.pone.0261768] [Medline: 35020727]
41. Developer agreement and policy. Developer Platform. URL: https://developer.twitter.com/en/developer-terms/agreement-and-policy [accessed 2021-03-06]
42. Xu K, Xie L, Yao K. Investigating LSTM for punctuation prediction. In: Proceedings of the 2016 10th International Symposium on Chinese Spoken Language Processing (ISCSLP). 2016 Presented at: 2016 10th International Symposium on Chinese Spoken Language Processing (ISCSLP); Oct 17-20, 2016; Tianjin, China. [doi: 10.1109/iscslp.2016.7918492]
43. Gulli A, Kapoor A, Pal S. Deep Learning with TensorFlow 2 and Keras Regression, ConvNets, GANs, RNNs, NLP, and More with TensorFlow 2 and the Keras API, 2nd Edition. Birmingham, United Kingdom: Packt Publishing; 2019.
44. Soliman AB, Eissa K, El-Beltagy SR. AraVec: a set of Arabic word embedding models for use in Arabic NLP. Procedia Comput Sci 2017;117:256-265. [doi: 10.1016/j.procs.2017.10.117]
45. Grave E, Bojanowski P, Gupta P, Joulin A, Mikolov T. Learning word vectors for 157 languages. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018). 2018 Presented at: Eleventh International Conference on Language Resources and Evaluation (LREC 2018); May 7-12, 2018; Miyazaki, Japan.
46. Fouad MM, Mahany A, Aljohani N, Abbasi RA, Hassan S. ArWordVec: efficient word embedding models for Arabic tweets. Soft Comput 2019 Jun 26;24(11):8061-8068. [doi: 10.1007/s00500-019-04153-6]
47. Farha I, Magdy W. Mazajak: an online Arabic sentiment analyser. In: Proceedings of the Fourth Arabic Natural Language Processing Workshop. 2019 Presented at: Fourth Arabic Natural Language Processing Workshop; Jul 28-Aug 2, 2019; Florence, Italy. [doi: 10.18653/v1/w19-4621]
48. Devlin J, Chang M, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). 2019 Presented at: 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers); Jun 2-7, 2019; Minneapolis, Minnesota. [doi: 10.18653/v1/N19-1423]
49. Shi S, Tang Z, Chu X, Liu C, Wang W, Li B. A quantitative survey of communication optimizations in distributed deep learning. IEEE Network 2021 May;35(3):230-237. [doi: 10.1109/mnet.011.2000530]
50. Safaya A, Abdullatif M, Yuret D. KUISAIL at SemEval-2020 Task 12: BERT-CNN for Offensive Speech Identification in Social Media. In: Proceedings of the Fourteenth Workshop on Semantic Evaluation. 2020 Presented at: Proceedings of the Fourteenth Workshop on Semantic Evaluation; Dec 12-13, 2020; Online. [doi: 10.18653/v1/2020.semeval-1.271]
51. Antoun W, Baly F, Hajj H. AraBERT: transformer-based model for Arabic language understanding. In: Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection. 2020 Presented at: Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection; May, 2020; Marseille, France.
52. Hugging Face. URL: https://huggingface.co/aubmindlab/bert-large-arabertv2 [accessed 2021-06-15]
53. Faraj D, Faraj D, Abdullah M. SarcasmDet at sarcasm detection task 2021 in Arabic using AraBERT pretrained model. In: Proceedings of the Sixth Arabic Natural Language Processing Workshop. 2021 Presented at: Proceedings of the Sixth Arabic Natural Language Processing Workshop; Apr 19, 2021; Kyiv, Ukraine (Virtual).
54. Darwish K, Mubarak H. Farasa: a new fast and accurate Arabic word segmenter. In: Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16). 2016 Presented at: Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16); May 2016; Portorož, Slovenia.
55. Abdelali A, Hassan S, Mubarak H, Darwish K, Samih Y. Pre-training BERT on Arabic tweets: practical considerations. arXiv 2021.
56. Abdul-Mageed M, Elmadany A, Nagoudi E. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). 2011 Presented at: The 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing; Aug, 2021; Online. [doi: 10.18653/v1/2021.acl-long.551]
57. Abu FI, Magdy W. Benchmarking transformer-based language models for Arabic sentiment and sarcasm detection. In: Proceedings of the Sixth Arabic Natural Language Processing Workshop. 2021 Presented at: Proceedings of the Sixth Arabic Natural Language Processing Workshop; Apr, 2021; Kyiv, Ukraine (Virtual).
58. Zhenyan L, Dan M, Weiping W, Chunxia Z. A supervised parameter estimation method of LDA. In: Proceedings of the Asia-Pacific Web Conference. 2015 Presented at: Asia-Pacific Web Conference; 2015; Guangzhou, China. [doi: https://doi.org/10.1007/978-3-319-25255-1_33]

59. Li H, Caragea D, Li X, Caragea C. Comparison of word embeddings and sentence encodings as generalized representations for crisis Tweet classification tasks. In: Proceedings of the ISCRAM Asian Pacific 2018 Conference. 2018 Presented at: ISCRAM Asian Pacific 2018 Conference; Nov, 2018; Wellington, New Zealand.
60. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychol Bull 1968 Oct;70(4):213-220. [doi: 10.1037/h0026256] [Medline: 19673146]
61. The measurement of interrater agreement. In: Statistical Methods for Rates and Proportions. Hoboken, New Jersey, United States: Wiley; 2003.
62. Nastasi A, Bryant T, Canner JK, Dredze M, Camp MS, Nagarajan N. Breast cancer screening and social media: a content analysis of evidence use and guideline opinions on twitter. J Cancer Educ 2018 Jun;33(3):695-702. [doi: 10.1007/s13187-017-1168-9] [Medline: 28097527]
63. Yamaguchi T, Shimizu J, Oya Y, Horio Y, Hida T. Drug-induced liver injury in a patient with nonsmall cell lung cancer after the self-administration of fenbendazole based on social media information. Case Rep Oncol 2021 Jun 17;14(2):886-891 [FREE Full text] [doi: 10.1159/000516276] [Medline: 34248555]
64. Powers D. J Mach Learn Technol 2011;2(1):37-63 [FREE Full text]
65. Hand D, Christen P. A note on using the F-measure for evaluating record linkage algorithms. Stat Comput 2017 Apr 19;28(3):539-547. [doi: 10.1007/s11222-017-9746-6]

Abbreviations

AraBERT: Transformer-based Model for Arabic Language Understanding
AraBERTv0.2: Transformer-based Model for Arabic Language Understanding version 0.2
AraBERTv2: Transformer-based Model for Arabic Language Understanding version 2
BERT: bidirectional encoder representations from transformers
BLSTM: bidirectional long short-term memory
DL: deep learning
LSTM: long short-term memory
ML: machine learning
NLP: natural language processing
RF: random forest
SM: social media
SVM: support vector machine
WHO: World Health Organization

Edited by A Mavragani; submitted 09.11.21; peer-reviewed by S Wei, A Rovetta, X Zhou; comments to author 08.02.22; revised version received 04.04.22; accepted 21.04.22; published 29.06.22.

Please cite as:
Albalawi Y, Nikolov NS, Buckley J
Pretrained Transformer Language Models Versus Pretrained Word Embeddings for the Detection of Accurate Health Information on Arabic Social Media: Comparative Study
JMIR Form Res 2022;6(6):e34834
URL: https://formative.jmir.org/2022/6/e34834
doi:10.2196/34834
PMID:35767322

Original Paper

Using Twitter Data for Cohort Studies of Drug Safety in Pregnancy: Proof-of-concept With β-Blockers

Ari Z Klein[1], PhD; Karen O'Connor[1], MS; Lisa D Levine[2], MD, MSCE; Graciela Gonzalez-Hernandez[3], PhD

[1]Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

[2]Department of Obstetrics and Gynecology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

[3]Department of Computational Biomedicine, Cedars-Sinai Medical Center, Los Angeles, CA, United States

Corresponding Author:
Ari Z Klein, PhD
Department of Biostatistics, Epidemiology, and Informatics
Perelman School of Medicine
University of Pennsylvania
Blockley Hall, 4th floor
423 Guardian Dr
Philadelphia, PA, 19014
United States
Phone: 1 310 423 3521
Email: ariklein@pennmedicine.upenn.edu

Abstract

Background: Despite the fact that medication is taken during more than 90% of pregnancies, the fetal risk for most medications is unknown, and the majority of medications have no data regarding safety in pregnancy.

Objective: Using β-blockers as a proof-of-concept, the primary objective of this study was to assess the utility of Twitter data for a cohort study design—in particular, whether we could identify (1) Twitter users who have posted tweets reporting that they took medication during pregnancy and (2) their associated pregnancy outcomes.

Methods: We searched for mentions of β-blockers in 2.75 billion tweets posted by 415,690 users who announced their pregnancy on Twitter. We manually reviewed the matching tweets to first determine if the user actually took the β-blocker mentioned in the tweet. Then, to help determine if the β-blocker was taken during pregnancy, we used the time stamp of the tweet reporting intake and drew upon an automated natural language processing (NLP) tool that estimates the date of the user's prenatal time period. For users who posted tweets indicating that they took or may have taken the β-blocker during pregnancy, we drew upon additional NLP tools to help identify tweets that report their pregnancy outcomes. Adverse pregnancy outcomes included miscarriage, stillbirth, birth defects, preterm birth (<37 weeks gestation), low birth weight (<5 pounds and 8 ounces at delivery), and neonatal intensive care unit (NICU) admission. Normal pregnancy outcomes included gestational age ≥37 weeks and birth weight ≥5 pounds and 8 ounces.

Results: We retrieved 5114 tweets, posted by 2339 users, that mention a β-blocker, and manually identified 2332 (45.6%) tweets, posted by 1195 (51.1%) of the users, that self-report taking the β-blocker. We were able to estimate the date of the prenatal time period for 356 pregnancies among 334 (27.9%) of these 1195 users. Among these 356 pregnancies, we identified 257 (72.2%) during which the β-blocker was or may have been taken. We manually verified an adverse pregnancy outcome—preterm birth, NICU admission, low birth weight, birth defects, or miscarriage—for 38 (14.8%) of these 257 pregnancies. We manually verified a gestational age ≥37 weeks for 198 (90.4%) and a birth weight ≥5 pounds and 8 ounces for 50 (22.8%) of the 219 pregnancies for which we did not identify an adverse pregnancy outcome.

Conclusions: Our ability to detect pregnancy outcomes for Twitter users who posted tweets reporting that they took or may have taken a β-blocker during pregnancy suggests that Twitter can be a complementary resource for cohort studies of drug safety in pregnancy.

(JMIR Form Res 2022;6(6):e36771) doi:10.2196/36771

KEYWORDS

natural language processing; social media; data mining; pregnancy; pharmacoepidemiology

Introduction

Prescription or over-the-counter medication is taken during more than 90% of pregnancies [1]. Despite the widespread use of medication during pregnancy, the fetal risk for most medications approved by the US Food and Drug Administration is unknown, and the majority of approved medications have no data regarding safety in pregnancy [2]. Given that Twitter has become a popular source of data on health conditions [3], it should be explored for evaluating drug safety in pregnancy, especially since 42% of people aged 18-29 years and 27% of people aged 30-49 years in the United States use Twitter [4]. Our prior work [5] used Twitter data in a case-control study that involved identifying users who reported a birth defect outcome (cases) [6] and users who did not (controls), and then searching their tweets for reports of medication exposure during pregnancy. Twitter data has not been assessed, however, for its utility in a cohort study design, which would involve identifying pregnancy outcomes for users who have reported taking medication during pregnancy.

Using β-blockers as a proof-of-concept, the primary objective of this study was to assess whether we could identify (1) Twitter users who have posted tweets reporting that they took medication during pregnancy and (2) their associated pregnancy outcomes, including miscarriage, stillbirth, birth defects, preterm birth (<37 weeks gestation), low birth weight (<5 pounds and 8 ounces at delivery), and neonatal intensive care unit (NICU) admission. We chose β-blockers as an example because cardiovascular disease is the leading cause of pregnancy-related deaths in the United States [7] and β-blockers are the most common type of medication for treating cardiac conditions during pregnancy [8]. Meanwhile, data on the safety of maternal β-blocker exposure are inconsistent; some studies report associations with low birth weight, preterm birth, perinatal mortality, or birth defects [9-17], while others do not [18-24].

Methods

Ethical Considerations

The Twitter data used in this study were collected and analyzed in accordance with the Twitter Terms of Service. The Institutional Review Board of the University of Pennsylvania reviewed this study and deemed it exempt human subjects research under 45 CFR §46.101(b)(4) for publicly available data sources (protocol# 828972). Although the tweets presented in this paper were public at the time of this study, we have slightly modified them, including removing usernames and URLs and redacting names, to help deidentify the users.

Medication Intake

We searched for mentions of β-blockers and their lexical variants (eg, misspellings) [25] in 2.75 billion tweets posted by 415,690 users who announced their pregnancy on Twitter [26]. Table 1 provides the β-blocker keywords and their lexical variants. We used annotation guidelines [27] to manually distinguish tweets reporting that the user actually took the β-blocker. If the tweet reported intake but did not explicitly indicate that the intake occurred during pregnancy, we used the time stamp of the tweet and drew upon an automated natural language processing (NLP) tool [28] that estimates the date of the user's prenatal time period. We also identified reports of taking a β-blocker that occurred before or after pregnancy, assuming that, if there was no evidence in the tweet that the user stopped taking it before pregnancy or started taking it after pregnancy, the user may have been taking it during pregnancy. We excluded users for whom we could not estimate the date of their prenatal time period.

Table 1. Keywords and their lexical variants used to search for tweets that mention β-blockers.

Keyword	Lexical variants
Acebutolol	N/A[a]
Atenolol	Atenelol, atenonol, atenanol, antenolol, atenol, atenolo, atenalol, antenenol, atentol, atenenol, attenalol, atenlol, attenolol, altenolol
Beta blocker	Beta-blocker, b blocker, b-blocker, beta blockers, beta-blockers, b blockers, b-blockers, betablocker, bblocker, betablockers, bblockers
Carvedilol	Carvidolol
Coreg	N/A
Corgard	N/A
Inderal	Inderall, inderol
Labetalol	Labetolol
Lopressor	N/A
Metoprolol	Metopolol, metropolol, metorolol, metroprolol, metaprolol, metoporol, metprolol, metotoprolol, metropolo, metroplol, meteprolol, metoprol, metroporol, metoprolo
Nadolol	Nadalol
Normodyne	N/A
Propranolol	Propananol, propanonol, proprapanol, propranonol, proranolol, propanolol, propranalol, propranonol, propanalol, propronolol
Sectral	N/A
Trandate	N/A
Tenormin	N/A
Toprol	Toprol, toprolol, topral, toperol, tropol, toporal, toporol, toporolol

[a]N/A: not applicable.

Pregnancy Outcomes

For users who posted tweets indicating that they took or may have taken the β-blocker during pregnancy, we drew upon automated NLP tools [29,30] to help identify tweets that self-report an associated pregnancy outcome, including miscarriage, stillbirth, birth defects, preterm birth, low birth weight, and NICU admission. To assess a potential reporting bias, we drew upon an automated NLP tool [31] that detects tweets reporting a gestational age ≥37 weeks (indicates the lack of miscarriage and preterm birth) or a birth weight ≥5 pounds and 8 ounces (indicates the lack of low birth weight, miscarriage, and stillbirth). If we did not automatically detect a tweet explicitly reporting a gestational age ≥37 weeks, we manually analyzed tweets posted during this time for evidence that the user was still pregnant.

Covariates

Two important potential confounders when evaluating drug safety in pregnancy are maternal age and indication for use. To help identify maternal age, we deployed an automated NLP tool [32] that identifies tweets self-reporting the exact age of the user at the time the tweet was posted. Then, we used the date of the user's prenatal time period to determine the user's age during pregnancy. To identify an indication for use, we manually reviewed the tweets reporting intake of a β-blocker posted by users who took or may have taken the β-blocker during pregnancy.

Results

Excluding retweets, we retrieved 5114 tweets, posted by 2339 users, that mention a β-blocker, and manually identified 2332 (45.6%) tweets, posted by 1195 (51.1%) of the users, that self-report taking the β-blocker. We were able to estimate the date of the prenatal time period for 334 (27.9%) of the 1195 users. Because some users' collection of tweets span several years and include multiple pregnancies, we identified 356 pregnancies among these 334 users. Among these 356 pregnancies, we found evidence that a β-blocker was taken during 58 (16.3%) of them and may have been taken during 199 (55.9%) of them. Table 2 presents examples of two users' tweets. User 1 reported on January 25, 2020, that the baby's due date was in 100 days, so our automated tool [28] estimated that pregnancy began on July 29, 2019, and would end on May 4, 2020. On April 16, 2020, User 1 explicitly reported taking Propranolol during pregnancy. User 1 reported that the baby was born premature on April 2, 2020—between 35 and 36 weeks gestation—with a low birth weight of 4 pounds and 12 ounces, and was admitted to the NICU. User 2 reported being 37 weeks pregnant on June 1, 2020, so our automated tool [28] estimated that pregnancy began on September 16, 2019, and would end on June 22, 2020. Whereas User 1 explicitly reported taking a β-blocker during pregnancy, for User 2, we used the time stamp of March 26, 2020, to infer that the intake was during pregnancy. User 2 reported on June 11, 2020, that the baby was

born—between 38 and 39 weeks gestation—and weighed 7 pounds and 5 ounces at birth.

We manually verified an adverse pregnancy outcome—preterm birth, NICU admission, low birth weight, birth defects, or miscarriage—for 38 (14.8%) of the 257 pregnancies during which a β-blocker was or may have been taken. Table 3 presents the adverse pregnancy outcomes among these 257 pregnancies. We manually verified a gestational age ≥37 weeks for 198 (90.4%) and a birth weight ≥5 pounds and 8 ounces for 50 (22.8%) of the 219 pregnancies for which we did not identify an adverse pregnancy outcome. We identified maternal age for 222 (86.4%) of the 257 pregnancies during which a β-blocker was or may have been taken. Table 3 includes the mean age per adverse pregnancy outcome. We identified an indication for taking the β-blocker for 197 (76.7%) of these 257 pregnancies—for example, tachycardia, hypertension, anxiety, and migraines.

Table 2. Sample tweets used to determine exposure to β-blockers during pregnancy and associated pregnancy outcomes.

User and tweet	Time stamp	Pregnancy start	Pregnancy end
1		2019-07-29	2020-05-04
exactly 100 days til my due date!	2020-01-25		
I was on Propranolol during my pregnancy and I had the CRAZIEST dreams I swear	2020-04-16		
Officially introducing [name], born April 2nd, 2020. 4lbs 12oz, 18”. She’s in NICU due to being premature, but she’s doing well!	2020-04-03		
2		2019-09-16	2020-06-22
5yo called me fat after I told 2.5yo I was too large to fit between their seats because of the baby. #37weekspregnant	2020-06-01		
I saw the MFM and cardiologist last week. It was determined my cardiomyopathy is manageable and I was put on a beta blocker	2020-03-26		
Introducing [name] 7lbs 5oz 20” long Csection went really well. We can’t wait until the big boys get to meet him	2020-06-11		

Table 3. Self-reported adverse pregnancy outcomes for Twitter users who took or may have taken a β-blocker during pregnancy (N=257).

Pregnancy outcome	Number of users, (%)	Sample tweet	Mean age
Preterm birth	23 (8.9%)	[name] came at 35 weeks. My baby is small, even for a preemie.	29 (n=20)
Neonatal intensive care unit admission	12 (4.7%)	Our sweet girl has been in the NICU these past few days. She's doing better everyday and we're really hoping she gets to go home soon.	27 (n=10)
Low birth weight	9 (3.5%)	Officially introducing [name], born April 2nd, 2020 at 11:01am. 4lbs 12oz, 18 inches.	27 (n=9)
Birth defect	4 (1.6%)	My son was also born with Craniosynostosis (Sagittal). He’s now 4 and wears his ‘wiggly’ line with pride	27 (n=3)
Miscarriage	1 (0.4%)	One of the worst parts of #miscarriage is ur 1st period afterwards. It’s so definitive, so confirming that it’s over #babyloss	45 (n=1)
Stillbirth	0 (0%)	N/A[a]	N/A
Composite[b]	38 (14.8%)	N/A	28 (n=33)

[a]N/A: not applicable.

[b]Multiple adverse pregnancy outcomes were identified for some pregnancies, so the number of composite adverse pregnancy outcomes is less than the sum of the individual adverse pregnancy outcomes.

Discussion

Principal Findings

Our ability to detect pregnancy outcomes for Twitter users who posted tweets reporting that they took or may have taken a β-blocker during pregnancy suggests more generally that Twitter could be a complementary resource for cohort studies of drug safety in pregnancy. Additionally, our ability to identify both the maternal age and indication for taking a β-blocker for many of the users demonstrates that Twitter data would even allow such studies to account for the effect of these two important potential confounders. This study suggests that Twitter data may be particularly valuable for assessing associations with preterm birth, given both the volume of its reports on Twitter and our finding that preterm birth is largely unaffected by a potential reporting bias; that is, we detected a gestational age ≥37 weeks for 198 (90.4%) of the 219 pregnancies for which we did not identify an adverse pregnancy outcome.

Limitations

Low birth weight may be affected by a potential reporting bias, given that we detected a birth weight ≥5 pounds and 8 ounces for only 50 (22.8%) of these 219 pregnancies. Although the rate of miscarriage in the United States is upward of more than 20% [33], our detection of miscarriage may be limited by a selection bias if users tend to announce their pregnancy on Twitter at a gestational age after which miscarriage infrequently occurs. Given our initial sample of 257 users, it is not surprising that we did not detect any reports of stillbirth, which has an incidence of <1% in the United States [34]. Nonetheless, our prior work [30] demonstrates that users do report stillbirth outcomes on Twitter, and our identification of users announcing their pregnancy on Twitter continues to grow in real time [26].

Conclusions

Given the widespread use of medication during pregnancy and the insufficient data on fetal risks, Twitter can be a complementary resource for cohort study designs.

Acknowledgments

The authors thank Ivan Flores for contributing to software applications, and Alexis Upshur for contributing to manually analyzing the Twitter data. This work was supported by the National Library of Medicine (R01LM011176).

Authors' Contributions

AZK contributed to collecting the Twitter data, manually analyzing the Twitter data, and writing the manuscript. KO contributed to manually analyzing the Twitter data and editing the manuscript. LDL designed the study, including the selection of β-blockers, pregnancy outcomes, and inclusion/exclusion criteria, and edited the manuscript. GGH conceptualized the use of Twitter data for studying medication exposure in pregnancy, guided the study, and edited the manuscript.

Conflicts of Interest

None declared.

References

1. Mitchell AA, Gilboa SM, Werler MM, Kelley KE, Louik C, Hernández-Díaz S, National Birth Defects Prevention Study. Medication use during pregnancy, with particular focus on prescription drugs: 1976-2008. Am J Obstet Gynecol 2011 Jul;205(1):51.e1-51.e8 [FREE Full text] [doi: 10.1016/j.ajog.2011.02.029] [Medline: 21514558]
2. Adam MP, Polifka JE, Friedman J. Evolving knowledge of the teratogenicity of medications in human pregnancy. Am J Med Genet C Semin Med Genet 2011 Aug 15;157C(3):175-182. [doi: 10.1002/ajmg.c.30313] [Medline: 21766440]
3. Mavragani A. Infodemiology and infoveillance: scoping review. J Med Internet Res 2020 Apr 28;22(4):e16206 [FREE Full text] [doi: 10.2196/16206] [Medline: 32310818]
4. Auxier B, Anderson M. Social media use in 2021. Pew Research Center. URL: https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/ [accessed 2021-08-02]
5. Golder S, Chiuve S, Weissenbacher D, Klein A, O'Connor K, Bland M, et al. Pharmacoepidemiologic evaluation of birth defects from health-related postings in social media during pregnancy. Drug Saf 2019 Mar 3;42(3):389-400 [FREE Full text] [doi: 10.1007/s40264-018-0731-6] [Medline: 30284214]
6. Klein AZ, Sarker A, Cai H, Weissenbacher D, Gonzalez-Hernandez G. Social media mining for birth defects research: A rule-based, bootstrapping approach to collecting data for rare health-related events on Twitter. J Biomed Inform 2018 Nov;87:68-78 [FREE Full text] [doi: 10.1016/j.jbi.2018.10.001] [Medline: 30292855]
7. Petersen EE, Davis NL, Goodman D, Cox S, Mayes N, Johnston E, et al. Vital Signs: Pregnancy-Related Deaths, United States, 2011-2015, and Strategies for Prevention, 13 States, 2013-2017. MMWR Morb Mortal Wkly Rep 2019 May 10;68(18):423-429 [FREE Full text] [doi: 10.15585/mmwr.mm6818e1] [Medline: 31071074]
8. Ruys TP, Maggioni A, Johnson MR, Sliwa K, Tavazzi L, Schwerzmann M, et al. Cardiac medication during pregnancy, data from the ROPAC. Int J Cardiol 2014 Nov 15;177(1):124-128. [doi: 10.1016/j.ijcard.2014.09.013] [Medline: 25499355]
9. Lydakis C, Lip G, Beevers M, Beevers D. Atenolol and fetal growth in pregnancies complicated by hypertension. Am J Hypertens 1999 Jun;12(6):541-547. [doi: 10.1016/s0895-7061(99)00031-x] [Medline: 10371362]
10. Magee L, Duley L. Oral beta-blockers for mild to moderate hypertension during pregnancy. Cochrane Database Syst Rev 2003(3):CD002863. [doi: 10.1002/14651858.CD002863] [Medline: 12917933]
11. Nakhai-Pour HR, Rey E, Bérard A. Antihypertensive medication use during pregnancy and the risk of major congenital malformations or small-for-gestational-age newborns. Birth Defects Res B Dev Reprod Toxicol 2010 Apr 31;89(2):147-154. [doi: 10.1002/bdrb.20238] [Medline: 20437474]
12. Meidahl Petersen K, Jimenez-Solem E, Andersen JT, Petersen M, Brødbæk K, Køber L, et al. β-Blocker treatment during pregnancy and adverse pregnancy outcomes: a nationwide population-based cohort study. BMJ Open 2012 Jul 19;2(4):e001185 [FREE Full text] [doi: 10.1136/bmjopen-2012-001185] [Medline: 22815467]

13. Yakoob MY, Bateman BT, Ho E, Hernandez-Diaz S, Franklin JM, Goodman JE, et al. The risk of congenital malformations associated with exposure to β-blockers early in pregnancy: a meta-analysis. Hypertension 2013 Aug;62(2):375-381 [FREE Full text] [doi: 10.1161/HYPERTENSIONAHA.111.00833] [Medline: 23753416]
14. Ersbøll AS, Hedegaard M, Søndergaard L, Ersbøll M, Johansen M. Treatment with oral beta-blockers during pregnancy complicated by maternal heart disease increases the risk of fetal growth restriction. BJOG 2014 Apr 09;121(5):618-626. [doi: 10.1111/1471-0528.12522] [Medline: 24400736]
15. Xie R, Guo Y, Krewski D, Mattison D, Walker M, Nerenberg K, et al. Beta-blockers increase the risk of being born small for gestational age or of being institutionalised during infancy. BJOG 2014 Aug 17;121(9):1090-1096. [doi: 10.1111/1471-0528.12678] [Medline: 24628701]
16. Tanaka K, Tanaka H, Kamiya C, Katsuragi S, Sawada M, Tsuritani M, et al. Beta-blockers and fetal growth restriction in pregnant women with cardiovascular disease. Circ J 2016;80(10):2221-2226. [doi: 10.1253/circj.cj-15-0617]
17. Duan L, Ng A, Chen W, Spencer HT, Lee M. Beta-blocker subtypes and risk of low birth weight in newborns. J Clin Hypertens (Greenwich) 2018 Nov 28;20(11):1603-1609 [FREE Full text] [doi: 10.1111/jch.13397] [Medline: 30267456]
18. Magee LA, Ornstein MP, von Dadelszen P. Fortnightly review: management of hypertension in pregnancy. BMJ 1999 May 15;318(7194):1332-1336 [FREE Full text] [doi: 10.1136/bmj.318.7194.1332] [Medline: 10323823]
19. Davis RL, Eastman D, McPhillips H, Raebel MA, Andrade SE, Smith D, et al. Risks of congenital malformations and perinatal events among infants exposed to calcium channel and beta-blockers during pregnancy. Pharmacoepidemiol Drug Saf 2011 Feb 15;20(2):138-145 [FREE Full text] [doi: 10.1002/pds.2068] [Medline: 21254284]
20. Duan L, Ng A, Chen W, Spencer HT, Nguyen J, Shen AY, et al. β-Blocker Exposure in Pregnancy and Risk of Fetal Cardiac Anomalies. JAMA Intern Med 2017 Jun 01;177(6):885-887 [FREE Full text] [doi: 10.1001/jamainternmed.2017.0608] [Medline: 28418448]
21. Bergman JEH, Lutke LR, Gans ROB, Addor M, Barisic I, Cavero-Carbonell C, et al. Beta-Blocker Use in Pregnancy and Risk of Specific Congenital Anomalies: A European Case-Malformed Control Study. Drug Saf 2018 Apr 11;41(4):415-427 [FREE Full text] [doi: 10.1007/s40264-017-0627-x] [Medline: 29230691]
22. Abalos E, Duley L, Steyn D, Gialdini C. Antihypertensive drug therapy for mild to moderate hypertension during pregnancy. Cochrane Database Syst Rev 2018 Oct 01;10:CD002252 [FREE Full text] [doi: 10.1002/14651858.CD002252.pub4] [Medline: 30277556]
23. Bateman BT, Heide-Jørgensen U, Einarsdóttir K, Engeland A, Furu K, Gissler M, et al. β-blocker use in pregnancy and the risk for congenital malformations: an international cohort study. Ann Intern Med 2018 Nov 20;169(10):665-673 [FREE Full text] [doi: 10.7326/M18-0338] [Medline: 30326014]
24. Wu Y, Yao J, Xu L, Chen M, Wan L. Risk of congenital malformations in offspring of women using β-blockers during early pregnancy: An updated meta-analysis of observational studies. Br J Clin Pharmacol 2021 Mar 27;87(3):806-815 [FREE Full text] [doi: 10.1111/bcp.14561] [Medline: 32978831]
25. Sarker A, Gonzalez-Hernandez G. An unsupervised and customizable misspelling generator for mining noisy health-related text sources. J Biomed Inform 2018 Dec;88:98-107 [FREE Full text] [doi: 10.1016/j.jbi.2018.11.007] [Medline: 30445220]
26. Sarker A, Chandrashekar P, Magge A, Cai H, Klein A, Gonzalez G. Discovering cohorts of pregnant women from social media for safety surveillance and analysis. J Med Internet Res 2017 Oct 30;19(10):e361 [FREE Full text] [doi: 10.2196/jmir.8164] [Medline: 29084707]
27. Klein A, Sarker A, Rouhizadeh M, O'Connor K, Gonzalez G. Detecting personal medication intake in Twitter: an annotated corpus and baseline classification system. In: Proceedings of the BioNLP 2017 Workshop. 2017 Presented at: BioNLP 2017; August 2017; Vancouver, Canada p. 136-142. [doi: 10.18653/v1/w17-2316]
28. Rouhizadeh M, Magge A, Klein A, Sarker A, Gonzalez G. A rule-based approach to determining pregnancy timeframe from contextual social media postings. In: DH '18: Proceedings of the 2018 International Conference on Digital Health. 2018 Presented at: Proceedings of the International Conference on Digital Health; April 2018; Lyon, France p. 23-26. [doi: 10.1145/3194658.3194679]
29. Klein AZ, Sarker A, Weissenbacher D, Gonzalez-Hernandez G. Towards scaling Twitter for digital epidemiology of birth defects. NPJ Digit Med 2019 Oct 01;2(1):96 [FREE Full text] [doi: 10.1038/s41746-019-0170-5] [Medline: 31583284]
30. Klein AZ, Cai H, Weissenbacher D, Levine LD, Gonzalez-Hernandez G. A natural language processing pipeline to advance the use of Twitter data for digital epidemiology of adverse pregnancy outcomes. J Biomed Inform 2020;112S:100076 [FREE Full text] [doi: 10.1016/j.yjbinx.2020.100076] [Medline: 34417007]
31. Klein AZ, Gebreyesus A, Gonzalez-Hernandez G. Automatically identifying comparator groups on Twitter for digital epidemiology of pregnancy outcomes. AMIA Jt Summits Transl Sci Proc 2020;2020:317-325 [FREE Full text] [Medline: 32477651]
32. Klein AZ, Magge A, Gonzalez-Hernandez G. ReportAGE: Automatically extracting the exact age of Twitter users based on self-reports in tweets. PLoS One 2022 Jan 25;17(1):e0262087 [FREE Full text] [doi: 10.1371/journal.pone.0262087] [Medline: 35077484]
33. Ammon Avalos L, Galindo C, Li D. A systematic review to calculate background miscarriage rates using life table analysis. Birth Defects Res A Clin Mol Teratol 2012 Jun 18;94(6):417-423. [doi: 10.1002/bdra.23014] [Medline: 22511535]

34. Hoyert D, Gregory E. Cause-of-death data from the fetal death file, 2015-2017. Natl Vital Stat Rep 2020 Apr;69(4):1-20 [FREE Full text] [Medline: 32510316]

Abbreviations

NICU: neonatal intensive care unit
NLP: natural language processing

Edited by A Mavragani; submitted 24.01.22; peer-reviewed by S Matsuda, V Foufi, M Beazely; comments to author 17.03.22; revised version received 27.04.22; accepted 06.06.22; published 30.06.22.

Please cite as:
Klein AZ, O'Connor K, Levine LD, Gonzalez-Hernandez G
Using Twitter Data for Cohort Studies of Drug Safety in Pregnancy: Proof-of-concept With β-Blockers
JMIR Form Res 2022;6(6):e36771
URL: https://formative.jmir.org/2022/6/e36771
doi:10.2196/36771
PMID:35771614

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org