JMIR Formative Research

Impact Factor (2023): 2.0
Volume 9 (2025) ISSN 2561-326X Editor in Chief: Amaryllis Mavragani, PhDc

Contents

Original Papers

Short Paper

Viewpoint

Real-Time Analytics and AI for Managing No-Show Appointments in Primary Health Care in the United Arab Emirates: Before-and-After Study

Yousif Mohamed AlSerkal[1*], BMed, MSc, MBBS, BAO; Naseem Mohamed Ibrahim[1*], MBBCH, Arab Board Speciality; Aisha Suhail Alsereidi[1*]; Mubaraka Ibrahim[1*]; Sudheer Kurakula[2*]; Sadaf Ahsan Naqvi[2*], MBBS, SQIL; Yasir Khan[2*]; Neema Preman Oottumadathil[2*]

[1]Emirates Health Services, Dubai, United Arab Emirates

[2]Oracle, The Edge Building, Al Falak Street, Dubai Internet City, Dubai, United Arab Emirates

[*]all authors contributed equally

Corresponding Author:
Sadaf Ahsan Naqvi, MBBS, SQIL
Oracle, The Edge Building, Al Falak Street, Dubai Internet City, Dubai, United Arab Emirates

Abstract

Background: Primary health care (PHC) services face operational challenges due to high patient volumes, leading to complex management needs. Patients access services through booked appointments and walk-in visits, with walk-in visits often facing longer waiting times. No-show appointments are significant contributors to inefficiency in PHC operations, which can lead to an estimated 3%-14% revenue loss, disrupt resource allocation, and negatively impact health care quality. Emirates Health Services (EHS) PHC centers handle over 140,000 visits monthly. Baseline data indicate a 21% no-show rate and an average patient wait time exceeding 16 minutes, necessitating an advanced scheduling and resource management system to enhance patient experiences and operational efficiency.

Objective: The objective of this study was to evaluate the impact of an artificial intelligence (AI)-driven solution that was integrated with an interactive real-time data dashboard on reducing no-show appointments and improving patient waiting times at the EHS PHCs.

Methods: This study introduced an innovative AI-based data application to enhance PHC efficiency. Leveraging our electronic health record system, we deployed an AI model with an 86% accuracy rate to predict no-shows by analyzing historical data and categorizing appointments based on no-show risk. The model was integrated with a real-time dashboard to monitor patient journeys and wait times. Clinic coordinators used the dashboard to proactively manage high-risk appointments and optimize resource allocation. The intervention was assessed through a before-and-after comparison of PHC appointment dynamics and wait times, analyzing data from 135,393 appointments (67,429 before implementation and 67,964 after implementation).

Results: Implementation of the AI-powered no-show prediction model resulted in a significant 50.7% reduction in no-show rates (P<.001). The odds ratio for no-shows after implementation was 0.43 (95% CI 0.42-0.45; P<.001), indicating a 57% reduction in the likelihood of no-shows. Additionally, patient wait times decreased by an average of 5.7 minutes overall (P<.001), with some PHCs achieving up to a 50% reduction in wait times.

Conclusions: This project demonstrates that integrating AI with a data analytics platform and an electronic health record systems can significantly improve operational efficiency and patient satisfaction in PHC settings. The AI model enabled daily assessments of wait times and allowed for real-time adjustments, such as reallocating patients to different clinicians, thus reducing wait times and optimizing resource use. These findings illustrate the transformative potential of AI and real-time data analytics in health care delivery.

(JMIR Form Res 2025;9:e64936) doi:10.2196/64936

KEYWORDS

electronic health record; EHR; artificial intelligence; AI; no-show appointments; real-time data; primary health care; risk prediction; clinic waiting time; operational efficiency

Introduction

Primary health care (PHC) centers are the initial contact point for health care and are considered the gateway to the health care system for a large proportion of the population [1,2]. PHC centers experience a high daily patient influx, making PHC operations complex and requiring efficient systems. While many patients book appointments in advance, others rely on walk-ins, often leading to longer wait times for older and pediatric patients who require quicker attention. However, a significant issue is no-show appointments, where patients fail to attend their scheduled appointments without prior cancellation.

No-show appointments create a significant burden by unnecessarily blocking health care resources. They can negatively impact the financial and operational planning of health care organizations, leading to an estimated revenue loss of 3% - 14% [3,4]. Various interventions have been proposed to address this issue, focusing on understanding patient behaviors and reducing no-show rates [5]. Many organizations use data analytics to predict no-show probabilities and devise interventions to improve clinical efficiencies [6-8]. Literature suggests that while machine learning models have been explored to predict no-show likelihoods [9], a gap exists in operationalizing these insights at a large scale, especially in settings like the Emirates Health Services (EHS), which manages a vast network of health care facilities.

Another key indicator of patient satisfaction is waiting time, which directly affects the quality of care and patient experience [10-12]. Numerous studies have demonstrated that reducing patient waiting times can significantly enhance patient satisfaction and overall service quality. These studies highlight that multiple interventions, such as resource optimization and system redesign, are necessary to improve operational efficiency. Identifying bottlenecks and implementing practical interventions can significantly reduce waiting times [13-17]. With the evolution of artificial intelligence (AI), organizations can now leverage real-time data analytics to develop predictive models and suggest proactive interventions.

While there are examples from various health care settings demonstrating the successful use of AI models and patient reminders to reduce no-shows [18,19], there is limited research on how such systems can be optimized specifically for large health care networks like the EHS. Several studies highlight the importance of real-time analytics in improving health care outcomes, allowing organizations to respond promptly to emerging challenges and optimize resource allocation [20-22]. Hence, this study aims to address this gap by providing an electronic health record (EHR)-driven, real-time analytics solution tailored to the operational needs of PHC centers.

The EHS is a vast health care network across 6 emirates in the United Arab Emirates, managing hospitals, clinics, specialized centers, and supporting services [23]. The PHC centers within the EHS handle over 140,000 visits each month, resulting in a high patient turnover. While this often leads to longer-than-expected wait times negatively impacting patient experience, some appointment slots remain underused due to an average no-shows rate of 21%. Baseline data show the average physician waiting time exceeds 16 minutes, and while resources are distributed based on center needs, occasional shortages contribute to operational delays.

Considering these challenges, our study aims to evaluate the impact of a real-time dashboard that projects operational metrics and integrates an AI model to predict no-show appointments in PHC centers at the EHS. By using this AI-enriched dashboard, clinic administrators can identify high-risk patients (risk of no-show) and proactively contact them to emphasize the importance of their scheduled visits. The primary goals of this evaluation were to assess if the dashboard improved operational efficiency, reduced no-show appointments, optimized resource allocation, and decreased patient waiting times.

Methods

Study Design

This project was conceived as an innovative data-driven intervention using our data intelligence platform seamlessly integrated with the EHR (Multimedia Appendix 1). Using real-time PHC data, we created a dashboard with an AI model to guide the operational workflow and to reduce clinic wait times and no-show appointments.

This was a before-and-after study that evaluated the impact of a real-time dashboard with an AI-powered no-show prediction model at EHS PHCs 3 months before and after implementation. By leveraging this dashboard, clinic coordinators could proactively manage PHC traffic flow, enhancing operational efficiency and reducing patient waiting times. To capture key metrics accurately, several data definitions were established to track patient flow within the PHC setting. Clinic waiting time was defined as the intervals between various points in a patient's journey, which were all recorded as time stamps in the EHR system. Specifically, this included:

- Nurse waiting time: time from registration to the nurse station.
- Physician waiting time: time from being ready to be seen by the physician to the actual consultation.
- Check-in to checkout time: total time from check-in to final checkout by the physician.
- Appointment no-show: defined as a patient missing a scheduled appointment without prior notice to the PHC.

For comprehensive analysis, all visit and attendance data were sourced from the EHR and the human resources attendance system, both of which were updated in real time to ensure data accuracy and relevance.

Research Road Map Intervention and Dashboard Implementation

Our research followed a structured path from data acquisition to intervention and evaluation. We first integrated patient and clinic staff data from the EHR into a real-time dashboard that incorporated an AI model trained on historical EHR data to predict no-show appointments. This model, along with other key PHC metrics—such as footfall, resource availability, and wait times—were displayed on the dashboard to support clinic coordinators in managing appointments proactively. The AI

model projected daily no-show risk, allowing coordinators to reach out to high-risk patients for confirmation or rescheduling. Additionally, real-time resource constraints were visible on the dashboard, enabling operational adjustments as needed. All users received training on the new dashboard and workflow, and continuous monitoring and evaluation processes were established to assess ongoing performance. Finally, we analyzed the results to measure the dashboard's impact on clinic operations (Multimedia Appendices 2 and 3).

AI Model

Our AI model was developed to predict no-show appointments using historical EHR data from EHS PHCs, incorporating 16 distinct features: 4 demographic, 3 on patient history, and 9 appointment-specific factors (Multimedia Appendix 4). We used a random forest classification technique to assess the model's effectiveness in handling high-dimensional data, achieving an accuracy rate of 86%. The model categorized appointments into a high, medium, or low risk of a no-show, based on probability thresholds of ≥90%, 89% - 80%, and 79% - 70%, respectively. To validate the AI model, we used bootstrapping to ensure robustness and to reduce overfitting.

Statistical Analysis and Experimental Validation

Data were analyzed using SPSS version 17 (IBM Corp) and SAS version 03.05 software (SAS Institute). Descriptive statistics were computed for all variables. Independent 2-tailed sample *t* tests were conducted to compare mean waiting times before and after the intervention. To analyze changes in appointment no-show rates, *z* tests were performed. Logistic regression was used to estimate odds ratios (OR) and determine the likelihood of a no-show after implementation. Statistical significance was set at an α level of .05, with 95% CI computed for all estimates. The statistical validation involved performing hypothesis tests to determine the significance of changes in waiting times and no-show rates.

Outcome Measures

The outcome measures were the mean difference in waiting times, which was calculated using the waiting times before and after implementation and the percent reduction in no-shows, which was calculated using the no-show rates before and after implementation.

Ethical Considerations

Ethical approval was obtained from the Dubai Research Ethics Committee at the Ministry of Health and Prevention, United Arab Emirates. (Approval Reference No. MOHAP/DXB-REC/S.O.O/No.136/2023). The study involved deidentified, retrospective data negating the need for individual patient consent. Patient compensation was not applicable for this study. Bias was minimized by using deidentified data and implementing objective data quality control measures.

Results

We implemented our project across all PHCs in the EHS in October 2022. To compare the outcomes of this tool, we analyzed the data 3 months before (July to September 2022) and 3 months after (November 2022, to January 2023) implementation, and the outcomes were grouped and labeled as before or after, respectively. Descriptive analytics revealed a total of 67,429 booked visits before implementation compared to 67,964 visits after implementation with a similar sex and nationality distribution between both periods (Multimedia Appendix 5). Our analysis showed that most of the visits were through walk-ins instead of booked appointments. There was a slight decrease in walk-in appointments between the two periods, but the percentage of these visits remained relatively stable (80.9% to 80.7%). There was a notable change in the number of no-show appointments, drastically reducing from 20.82% to 10.25% (Table 1).

Table . Visit frequencies and dynamics before and after program implementation in primary health care.

	Before implementation, n/N (%)	After implementation, n/N (%)
Walk-ins	285,700/353,129 (80.91)	284,015/351,979 (80.69)
Appointments	67,429/353,129 (19.09)	67,964/351,979 (19.31)
No-shows	14,038/67,429 (20.82)	6966/67,964 (10.25)

Using a *z* test, our analysis showed that the implementation of the no-show AI model led to a significant 50.7% reduction in no-show rates across the PHC department ($P<.001$). Binary logistic regression further assessed the dashboard's impact on reducing no-shows, yielding an OR of 0.43 (95% CI 0.42 - 0.45; $P<.001$). This OR being less than 1 indicated that the intervention effectively decreased no-show occurrences, reducing the likelihood of no-shows by 57% after implementation of the AI model (Table 2).

Table . Mean difference and impact analysis of appointment no-show rates before and after program implementation in primary health care centers.

Measurements	Values	*P* value
Difference in no-show rate after implementation, n/N (%)	50.7	<.001
Likelihood of no-shows after implementation, odds ratio (95% CI)	0.43 (0.42-0.45)	<.001

Table 3 reflects the different waiting times for the nurse assessment, physician assessment, and overall consultation time (Multimedia Appendix 6). We observed reduced average waiting times at all stations across the PHC journey. We found a

decrease in patient wait times by an average of 5.7 minutes (from 54 to 49 minutes; $P<.001$). Based on the average time saved per visit (5.7 minutes), a total of 387,394.8 minutes were saved in the 3 months after implementation. This is equivalent to approximately 6,456.6 hours saved during this period. The reduction in waiting time from check-in to checkout has been consistent since implementing this solution.

Table . Comparison of patient waiting times at different stations during patient visits before and after program implementation at primary health care centers.

Waiting times	Before implementation (min), mean (SE)	After implementation (min), mean (SE)	Difference following implementation (min)	*P* value
Nurse waiting time	8.4 (0.02)	7.2 (0.02)	1.2	<.001
Physician waiting time	16.3 (0.03)	15.0 (0.03)	1.3	<.001
Physician consultation to checkout time	29.3 (0.06)	26.2 (0.06)	3.1	<.001
Check-in to checkout time	54.2 (0.08)	48.5 (0.07)	5.7	<.001

Discussion

Principal Results and Comparison With Prior Work

The implementation of an AI-powered no-show prediction model, integrated with a real-time operational dashboard at PHC centers within the EHS, has demonstrated significant improvements in operational efficiency and patient care. Our main findings revealed a significant 50.7% reduction in no-show appointments *(P*<.001), and a 5.7-minute reduction in overall patient waiting times, highlighting the direct impact of our intervention on streamlining health care delivery. These results underlined the value of leveraging predictive analytics and real-time data in optimizing resource allocation, reducing manual processes, and enhancing the patient experience. Unlike other studies, where the emphasis has been to explore the factors and behaviors contributing to patient no-shows using machine learning and other techniques [24], we successfully operationalized our machine learning model and connected it to a closed loop workflow which helped reduce absenteeism.

A key achievement of this intervention was the substantial reduction in no-show appointments. By proactively contacting high-risk patients identified by the AI model, clinic administrators were able to significantly reduce missed appointments. This was a notable success, as it not only improved the efficiency of the appointments but also allowed walk-in patients to fill the available slots, further maximizing clinic capacity. Additionally, the use of real-time dashboards provided clinic managers with immediate insights into patient volumes, waiting times, and staff availability, enabling swift adjustments in resource allocation during peak hours. These operational efficiencies, particularly the reduction in waiting times at multiple stages of the patient journey (nurse assessment, physician consultation, and overall clinic visit), underscore the importance of real-time analytics in improving the flow of care. Our robust analysis reassured that the findings of this before and after implementation were significant and could be attributed to our intervention.

The impact of these factors has been previously studied in isolation. There are successful examples of AI models to predict no-show appointments, which have helped reduce the burden of these instances [25-27]. Our observations were in alignment with these studies. Like other studies, we have seen that connecting with patients before their scheduled appointment reduces their chances of missing it [28]. Several strategies, like text messages, patient portal reminders, and phone calls, have been examined, and it has been found that calling patients before their scheduled visit is effective for reducing no-shows [29].

Operationally, real-time dashboards and decision support tools for clinic managers are helpful for identifying operational bottlenecks. Numerous studies have examined the impact of real-time dashboards on reducing patient waiting times in health care settings. These studies suggest that real-time dashboards are crucial for improving patient flow and reducing waiting times across various health care settings, emphasizing their potential as valuable tools in optimizing health care delivery [30-33].

As recommended by other studies, using data to calculate and predict interday scheduling precisely is undeniably beneficial for large health care organizations [30,34]. Trend analysis helps estimate the expected number of unbooked versus booked visits and helps allocate resources accordingly. Many PHC centers reserve a proportion of their scheduling slots for walk-in visits. Since the EHS has many walk-in visits, we should follow a similar strategy and use predictive analysis to help improve our resource allocation. We must carefully design this strategy to reduce our patient waiting times. In the future, we could further enhance the outcomes achieved in this study. By scaling up the AI-model to other services and departments, we could increase the benefits of this program for the EHS. The flexibility and the scalability of our real-time data intelligence platform makes it a valuable tool to resolve operational challenges in our health care system.

Strengths

This study presents a novel approach by integrating an AI-driven predictive model with a real-time dashboard to manage no-show appointments and reduce waiting times in a PHC setting. Unlike previous studies, which primarily focus on predictive modeling without an operational framework, our approach goes a step further by embedding predictive insights into an actionable, real-time dashboard that enables proactive intervention. By operationalizing predictive analytics within a large-scale health care network, this study contributes a scalable and practical

solution to the current state of digital health interventions aimed at improving health care efficiency and patient outcomes.

The major strengths of our study include big data analysis, the use of case mix, the contextualization to the local population, crossfunctional analytics for the clinician-to-patient ratio, real-time analysis for immediate identification of prolonged waiting times, and AI-based modeling for no-show appointment prediction that is coupled with patient contact information. Our holistic approach to use an AI model and real-time insights for a guided operational management workflow is unique in its nature and it can be used by organizations with similar resources.

Limitations

Certain limitations of this study must be acknowledged. First, the mitigation workflow is outside of EHRs, so we cannot quantify the efforts accurately. It is an established fact that a reduction in no-show appointments result in improved financial outcomes [35,36], and we acknowledge that a detailed analysis of the financial implications of reducing no-shows was beyond the scope of this study. Future research could address this aspect to provide a more comprehensive understanding of the intervention's economic impact.

Conclusion

In conclusion, this project represents the first instance of the EHS using a real-time, AI-driven analytics platform to improve operational efficiency, reduce no-show appointments, and enhance patient experience. The success of our solution in reducing no-shows and waiting times highlights the importance of integrating advanced analytics into daily health care operations. By providing clinic administrators with real-time actionable data, this platform enables better decision-making, optimized resource use, and improved patient flow. The scalability of this application across other health care settings further emphasizes its potential to drive widespread improvements in health care delivery, making it a valuable tool for health care organizations looking to modernize their operations through data-driven solutions. As we continue to refine this system, we expect further reductions in waiting times and improvements in patient satisfaction.

Acknowledgments

We would like to acknowledge Benjamin Vinodson. The study was fully funded by the Emirates Health Services, with no external sources involved.

Data Availability

The data sets generated during and/or analyzed during this study are available from the corresponding author on reasonable request

Authors' Contributions

NMI was the principal investigator of this project. All authors contributed to the design, analysis, and manuscript creation according to the sequence of authorship assigned.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Data architecture diagram representing the layers of data transition from source of information to the dashboard.
[PPTX File, 718 KB - formative_v9i1e64936_app1.pptx]

Multimedia Appendix 2
Program workflow for using the real-time artificial intelligence–powered dashboard to manage no-shows and reduce waiting times in primary health care.
[PPTX File, 45 KB - formative_v9i1e64936_app2.pptx]

Multimedia Appendix 3
A screenshot of the data dashboard showing real-time analytics for primary health care visits in a day.
[PPTX File, 1025 KB - formative_v9i1e64936_app3.pptx]

Multimedia Appendix 4
Major predictors contributing to the no-show prediction artificial intelligence model from primary health care visit data.
[PPTX File, 139 KB - formative_v9i1e64936_app4.pptx]

Multimedia Appendix 5

Frequency distribution of booked visits before and after the artificial intelligence program implementation in primary health care centers (July 2022, to January 2023).
[PDF File, 84 KB - formative_v9i1e64936_app5.pdf]

Multimedia Appendix 6
Reduction in patient waiting times at different stages of care following artificial intelligence program implementation in primary health care centers.
[PPTX File, 48 KB - formative_v9i1e64936_app6.pptx]

References

1. Primary health care. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/primary-health-care [accessed 2024-10-09]
2. Leslie LK, Mehus CJ, Hawkins JD, et al. Primary health care: potential home for family-focused preventive interventions. Am J Prev Med 2016 Oct;51(4 Suppl 2):S106-S118. [doi: 10.1016/j.amepre.2016.05.014] [Medline: 27498167]
3. Guzek LM, Gentry SD, Golomb MR. The estimated cost of “no-shows” in an academic pediatric neurology clinic. Pediatr Neurol 2015 Feb;52(2):198-201. [doi: 10.1016/j.pediatrneurol.2014.10.020] [Medline: 25497123]
4. Moore CG, Wilson-Witherspoon P, Probst JC. Time and money: effects of no-shows at a family practice residency clinic. Fam Med 2001;33(7):522-527. [Medline: 11456244]
5. Harris SL, Samorani M. On selecting a probabilistic classifier for appointment no-show prediction. Decis Support Syst 2021 Mar;142:113472. [doi: 10.1016/j.dss.2020.113472]
6. Lin Q, Betancourt B, Goldstein BA, Steorts RC. Prediction of appointment no-shows using electronic health records. J Appl Stat 2020;47(7):1220-1234. [doi: 10.1080/02664763.2019.1672631] [Medline: 35707022]
7. Denney J, Coyne S, Rafiqi S. Machine learning predictions of no-show appointments in a primary care setting. SMU Data Sci Rev 2019;2(1) [FREE Full text]
8. Harvey HB, Liu C, Ai J, et al. Predicting no-shows in radiology using regression modeling of data available in the electronic medical record. J Am Coll Radiol 2017 Oct;14(10):1303-1309. [doi: 10.1016/j.jacr.2017.05.007] [Medline: 28673777]
9. Yang Y, Madanian S, Parry D. Enhancing health equity by predicting missed appointments in health care: machine learning study. JMIR Med Inform 2024 Jan 12;12:e48273. [doi: 10.2196/48273] [Medline: 38214974]
10. Dansky KH, Miles J. Patient satisfaction with ambulatory healthcare services: waiting time and filling time. Hosp Health Serv Adm 1997;42(2):165-177. [Medline: 10167452]
11. Huang XM. Patient attitude towards waiting in an outpatient clinic and its applications. Health Serv Manage Res 1994 Feb;7(1):2-8. [doi: 10.1177/095148489400700101] [Medline: 10133292]
12. Leddy KM, Kaldenberg DO, Becker BW. Timeliness in ambulatory care treatment. An examination of patient satisfaction and wait times in medical practices and outpatient test and treatment facilities. J Ambul Care Manage 2003;26(2):138-149. [doi: 10.1097/00004479-200304000-00006] [Medline: 12698928]
13. Michael M, Schaffer SD, Egan PL, Little BB, Pritchard PS. Improving wait times and patient satisfaction in primary care. J Healthc Qual 2013;35(2):50-59. [doi: 10.1111/jhq.12004] [Medline: 23480405]
14. Vinitsky A, David B, Michalik L, et al. Reducing wait time in a high-volume pediatric neuro-oncology clinic by optimizing process flow: a quality improvement project. Pediatr Qual Saf 2022;7(3):e557. [doi: 10.1097/pq9.0000000000000557] [Medline: 35720861]
15. Santibáñez P, Chow VS, French J, Puterman ML, Tyldesley S. Reducing patient wait times and improving resource utilization at British Columbia Cancer Agency’s ambulatory care unit through simulation. Health Care Manag Sci 2009 Dec;12(4):392-407. [doi: 10.1007/s10729-009-9103-1] [Medline: 20058528]
16. Le V, Wagar EA, Phipps RA, Del Guidice RE, Le H, Middleton LP. Improving patient experience of wait times and courtesy through electronic sign-in and notification in the phlebotomy clinic. Arch Pathol Lab Med 2020 Jun;144(6):769-775. [doi: 10.5858/arpa.2019-0139-OA] [Medline: 31755777]
17. Almomani I, AlSarheed A. Enhancing outpatient clinics management software by reducing patients’ waiting time. J Infect Public Health 2016;9(6):734-743. [doi: 10.1016/j.jiph.2016.09.005] [Medline: 27663518]
18. Saeed S, Somani N, Sharif F, Kazi AM. Evaluating the effectiveness of text messaging and phone call reminders to minimize no show at pediatric outpatient clinics in Pakistan: protocol for a mixed-methods study. JMIR Res Protoc 2018 Apr 10;7(4):e91. [doi: 10.2196/resprot.9294] [Medline: 29636321]
19. Horvath M, Levy J, L’Engle P, Carlson B, Ahmad A, Ferranti J. Impact of health portal enrollment with email reminders on adherence to clinic appointments: a pilot study. J Med Internet Res 2011 May 26;13(2):e41. [doi: 10.2196/jmir.1702] [Medline: 21616784]
20. Milosevic Z, Chen W, Berry A, Rabhi F. Real-time analytics. In: Big Data 2016:39-61. [doi: 10.1016/B978-0-12-805394-2.00002-7]
21. Wills MJ. Decisions through data: analytics in healthcare. J Healthc Manag 2014;59(4):254-262. [doi: 10.1097/00115514-201407000-00005] [Medline: 25154123]

22. Kaur J, Kulwinder Singh Mann D. AI based healthcare platform for real time, predictive and prescriptive analytics using reactive programming. J Phys Conf Ser 2018 Jan 3;933(1):012010. [doi: 10.1088/1742-6596/933/1/012010]
23. Emirates Health Services - UAE. URL: https://www.ehs.gov.ae/en/home [accessed 2024-12-20]
24. Salazar LHA, Parreira WD, Fernandes ADR, Leithardt VRQ. No-show in medical appointments with machine learning techniques: a systematic literature review. Inf 2022;13(11):507. [doi: 10.3390/info13110507]
25. AlMuhaideb S, Alswailem O, Alsubaie N, Ferwana I, Alnajem A. Prediction of hospital no-show appointments through artificial intelligence algorithms. Ann Saudi Med 2019;39(6):373-381. [doi: 10.5144/0256-4947.2019.373] [Medline: 31804138]
26. Goffman RM, Harris SL, May JH, et al. Modeling patient no-show history and predicting future outpatient appointment behavior in the veterans health administration. Mil Med 2017 May;182(5):e1708-e1714. [doi: 10.7205/MILMED-D-16-00345] [Medline: 29087915]
27. Dunstan J, Villena F, Hoyos JP, et al. Predicting no-show appointments in a pediatric hospital in Chile using machine learning. Health Care Manag Sci 2023 Jun;26(2):313-329. [doi: 10.1007/s10729-022-09626-z] [Medline: 36707485]
28. Liu C, Harvey HB, Jaworsky C, Shore MT, Guerrier CE, Pianykh O. Text message reminders reduce outpatient radiology no-shows but do not improve arrival punctuality. J Am Coll Radiol 2017 Aug;14(8):1049-1054. [doi: 10.1016/j.jacr.2017.04.016] [Medline: 28583321]
29. Childers RE, Laird A, Newman L, Keyashian K. The role of a nurse telephone call to prevent no-shows in endoscopy. Gastrointest Endosc 2016 Dec;84(6):1010-1017. [doi: 10.1016/j.gie.2016.05.052] [Medline: 27327847]
30. Cayirli T, Yang KK, Quek SA. A universal appointment rule in the presence of no‐shows and walk‐ins. Prod Oper Manag 2012 Jul;21(4):682-697. [doi: 10.1111/j.1937-5956.2011.01297.x]
31. Franklin A, Gantela S, Shifarraw S, et al. Dashboard visualizations: supporting real-time throughput decision-making. J Biomed Inform 2017 Jul;71:211-221. [doi: 10.1016/j.jbi.2017.05.024] [Medline: 28579532]
32. Marmor YN, Wasserkrug S, Zeltyn S, et al. Toward simulation-based real-time decision-support systems for emergency departments. Presented at: The 2009 Winter Simulation Conference (WSC); Dec 13-16, 2009; Austin, TX p. 2042-2053. [doi: 10.1109/WSC.2009.5429416]
33. Alaeddini A, Murty KG. DSS (decision support system) for allocating appointment times to calling patients at a medical facility. In: Murty KG, editor. Case Studies in Operations Research: Springer; 2015:83-109. [doi: 10.1007/978-1-4939-1007-6_5]
34. Wiesche L, Schacht M, Werners B. Strategies for interday appointment scheduling in primary care. Health Care Manag Sci 2017 Sep;20(3):403-418. [doi: 10.1007/s10729-016-9361-7] [Medline: 27000079]
35. Berg BP, Murr M, Chermak D, et al. Estimating the cost of no-shows and evaluating the effects of mitigation strategies. Med Decis Making 2013 Nov;33(8):976-985. [doi: 10.1177/0272989X13478194] [Medline: 23515215]
36. Kheirkhah P, Feng Q, Travis LM, Tavakoli-Tabasi S, Sharafkhaneh A. Prevalence, predictors and economic consequences of no-shows. BMC Health Serv Res 2016 Jan 14;16:13. [doi: 10.1186/s12913-015-1243-z] [Medline: 26769153]

Abbreviations

AI: artificial intelligence
EHR: electronic health record
EHS: Emirates Health Services
OR: odds ratio
PHC: primary health care

Edited by A Mavragani; submitted 12.08.24; peer-reviewed by A Ala, F Yang; revised version received 14.11.24; accepted 14.11.24; published 06.01.25.

Please cite as:
AlSerkal YM, Ibrahim NM, Alsereidi AS, Ibrahim M, Kurakula S, Naqvi SA, Khan Y, Oottumadathil NP
Real-Time Analytics and AI for Managing No-Show Appointments in Primary Health Care in the United Arab Emirates: Before-and-After Study
JMIR Form Res 2025;9:e64936
URL: https://formative.jmir.org/2025/1/e64936
doi:10.2196/64936

Effects of Smart Goggles Used at Bedtime on Objectively Measured Sleep and Self-Reported Anxiety, Stress, and Relaxation: Pre-Post Pilot Study

Sharon Danoff-Burg, PhD; Elie Gottlieb, PhD; Morgan A Weaver, BA; Kiara C Carmon, BA; Duvia Lara Ledesma, MA; Holly M Rus, PhD

Scientific Research, SleepScore Labs, 2175 Salk Avenue, Suite 150, Carlsbad, CA, United States

Corresponding Author:
Sharon Danoff-Burg, PhD
Scientific Research, SleepScore Labs, 2175 Salk Avenue, Suite 150, Carlsbad, CA, United States

Abstract

Background: Insufficient sleep is a problem affecting millions. Poor sleep can trigger or worsen anxiety; conversely, anxiety can lead to or exacerbate poor sleep. Advances in innovative consumer products designed to promote relaxation and support healthy sleep are emerging, and their effectiveness can be evaluated accurately using sleep measurement technologies in the home environment.

Objective: This pilot study examined the effects of smart goggles used at bedtime to deliver gentle, slow vibration to the eyes and temples. The study hypothesized that objective sleep, perceived sleep, self-reported stress, anxiety, relaxation, and sleepiness would improve after using the smart goggles.

Methods: A within-participants, pre-post study design was implemented. Healthy adults with subclinical threshold sleep problems (N=20) tracked their sleep nightly using a polysomnography-validated noncontact biomotion device and completed daily questionnaires over two phases: a 3-week baseline period and a 3-week intervention period. During the baseline period, participants followed their usual sleep routines at home. During the intervention period, participants used Therabody SmartGoggles in "Sleep" mode at bedtime. This mode, designed for relaxation, delivers a gentle eye and temple massage through the inflation of internal compartments to create a kneading sensation combined with vibrating motors. Each night, the participants completed questionnaires assessing relaxation, stress, anxiety, and sleepiness immediately before and after using the goggles. Daily morning questionnaires assessed perceived sleep, complementing the objective sleep data measured every night.

Results: Multilevel regression analysis of 676 nights of objective sleep parameters showed improvements during nights when the goggles were used compared to the baseline period. Key findings include sleep duration (increased by 12 minutes, P=.01); duration of deep sleep (increased by 6 minutes, P=.002); proportion of deep sleep (7% relative increase, P=.02); BodyScore, an age- and gender-normalized measure of deep sleep (4% increase, P=.002); number of nighttime awakenings (7% decrease, P=.02); total time awake after sleep onset (reduced by 6 minutes, P=.047); and SleepScore, a measure of overall sleep quality (3% increase, P=.02). Questionnaire responses showed that compared to baseline, participants felt they had better sleep quality (P<.001) and woke feeling more well-rested (P<.001). Additionally, participants reported feeling sleepier, less stressed, less anxious, and more relaxed (all P values <.05) immediately after using the goggles each night, compared to immediately before use. A standardized inventory administered before and after the 3-week intervention period indicated reduced anxiety (P=.03), confirming the nightly analysis.

Conclusions: The use of smart goggles at bedtime significantly improved objectively measured sleep metrics and perceived sleep quality. Further, participants reported increased feelings of relaxation along with reduced stress and anxiety. Future research expanding on this pilot study is warranted to confirm and expand on the preliminary evidence presented in this brief report.

(JMIR Form Res 2025;9:e58461) doi:10.2196/58461

KEYWORDS

relaxation; stress; anxiety; sleep; health technology; intervention

Introduction

Insufficient sleep affects approximately one-third of the population [1] and is associated with adverse health outcomes and impaired performance [2]. Poor sleep can trigger or worsen anxiety, while conversely, anxiety can lead to or exacerbate poor sleep [3]. Technological advances in unobtrusive sleep measurement enable intervention studies to be conducted in the comfort of research participants' own bedrooms, providing ecologically valid results while capturing accurate objective

data [4,5]. Concomitantly, the development of innovative consumer products designed to promote relaxation and support healthy sleep is emerging. Their effectiveness can be evaluated in field studies using ambulatory measurement technologies [6].

The aim of this research was to examine the effects of smart goggles used at bedtime to deliver gentle vibration to the eyes and temples, on sleep as well as on perceived stress, anxiety, and relaxation. Although a variety of evidence-based relaxation techniques [7] and sleep-enhancing products [8] already exist, devices such as smart goggles may appeal to individuals who wish to use such technological tools as an option within their repertoire of strategies for winding down at bedtime. Previous research suggests that vibration can increase relaxation [9], induce drowsiness [10], and may be a useful nonpharmacological intervention for poor sleep [11-13]. For example, a preliminary study documented the use of vibration in improving objectively measured sleep outcomes in people with mild to moderate symptoms of insomnia [11]. In this pilot study on adults with subclinical threshold sleep problems, we hypothesized that objective and perceived sleep outcomes as well as self-reported stress, anxiety, relaxation, and sleepiness would improve after using smart goggles delivering slow vibrations at bedtime.

Methods

Participants

Invitations to complete an eligibility questionnaire for a study testing smart goggles were emailed to registrants in a large database of individuals interested in sleep research and using SleepScore technology. Eligible respondents were invited to enroll based on the following selection criteria: difficulty falling or staying asleep, no history of sleep disorders, absence of other medical issues affecting sleep, substance use that could affect sleep, and no lifestyle issues such as shift work that might influence their sleep. The study included adults (N=20) with subclinical threshold sleep problems who were willing to track their sleep and use the smart goggles as instructed.

Design and Procedures

A within-participants, pre-post study design was implemented. The participants were aware that the intervention had the potential to affect their sleep. Following a 3-week baseline period during which participants measured their sleep data at home without any intervention, they used the Therabody SmartGoggles (Therabody Inc, Los Angeles, CA) at bedtime for 15 minutes (within 30 minutes of their intended sleep time) over a 3-week intervention period. The participants were instructed to use “Sleep” mode, designed for relaxation and inducing sleepiness. This mode delivers slow and gentle massage to the eye and temple areas by the inflation of internal compartments to create a kneading sensation and vibrating motors. Additionally, two other modes (SmartRelax and Focus) are available; however, since these provide different experiences, they were not used in this study.

During the entire 6-week study, nightly objective sleep measurements were collected and participants completed online questionnaires each morning and evening. Data collection was synchronized across all participants to account for weekday or weekend variation.

Measurement

Objective sleep data were collected with SleepScore Max (Consumer Sleep Solutions LLC, Carlsbad, CA), a noncontact monitoring device that uses respiratory signal and motion sensing to detect sleep. The device is placed next to the bed and controlled using a companion app. It uses ultra-low power radiofrequency waves to monitor body movement and respiration patterns when in bed; the measurement is unaffected by bedding or nightwear. If a partner is present, only the sleep of the individual closest to the device is measured. The device captures high-resolution magnitude and duration data of gross movements, micromovements, and full breathing cycles, which are transformed into 30-second epoch sleep stage data (wake, light, deep, rapid eye movement [REM]) using proprietary algorithms. Studies have shown good agreement with gold-standard polysomnography [14,15], exceeding the accuracy typically reported for actigraphy-based devices [16].

Using the 30-second epoch data, standard sleep metrics were calculated. In addition, 3 SleepScore technology proprietary sleep metrics reflecting sleep quality, all ranging from 0 to 100 and normalized for age and gender using reference values from the meta-analysis of quantitative sleep parameters by Ohayon and colleagues [17], were calculated:

- *SleepScore* is an overall sleep quality metric that includes objectively measured total sleep time, sleep onset latency, and sleep stage durations.
- *BodyScore* reflects the age- and gender-normalized amount of deep (non–rapid eye movement stage 3 [NREM-3]) sleep.
- *MindScore* reflects the age- and gender-normalized amount of REM sleep.

Self-reported data were collected daily, across the entire study via 100-point visual analog scales. Morning assessments measured perceived sleep quality and feeling well-rested upon waking. At night, the 100-point visual analog scales assessed relaxation, stress, anxiety, and sleepiness before and after goggle use. The construct validity and discriminate sensitivity of visual analog scales to assess perceived stress and related constructs have been documented [18]. Participants completed these scales immediately before using the goggles at bedtime and after using the goggles for 15 minutes, and then went to sleep. A 6-item version of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI) [19] was administered once before and once after the intervention period.

Statistical Analyses

Nightly objective sleep data and daily self-reported data were analyzed using multilevel regression with random intercept, accounting for nights nested within participants, comparing nights during the baseline period to nights during the intervention period for each outcome. The regression model was $Sleepmeasure_{ij} = \beta_0 + \beta_1 * TestPeriod_{ij} + u_{0j} + e_{ij}$; TestPeriod, coded as 0 for observations during baseline and 1 for nights during the intervention period. Similarly, analysis of the nightly self-reported data used the same model to compare pre- and

post-goggle use. A 2-tailed paired-samples *t* test was used to analyze changes in the 6-item STAI scores.

Discrepancies in sample sizes (N=20 for objective and self-reported sleep, n=17 for 6-item STAI) were due to incomplete data sources. Participants tracked their sleep at home and, at times, were not fully compliant with the use of measurement tools or the completion of online surveys. This is common in field research collecting longitudinal and daily assessment data. All reported results reflect the largest available sample for each set of analyses.

Ethical Considerations

The study was approved by Sterling Institutional Review Board (ID 11012), and all procedures were conducted in accordance with the ethical standards of the Declaration of Helsinki. All participants provided informed consent using a secure online platform to review, electronically sign, and return a copy of the document to the research coordinator. They were informed that the study was voluntary, and of their right to withdraw at any time. Both objective and self-reported study data were deidentified prior to analysis and accessible only to members of the research team. As compensation, participants kept the smart goggles and sleep measurement device used during the study.

Results

Of the 20 participants, 40% (n=8) were women, and the age range was 26-75 (mean 50.41, SD 13.12) years. Further demographic details are provided in Table 1.

Table . Demographic information of participants at baseline (N=20).

Demographics	Values
Age (years)	
Mean (SD)	50.41 (13.12)
Range	26 - 75
Gender, n (%)	
Men	8 (40)
Women	12 (60)
Other identities	0 (0)
Race/Ethnicity, n (%)	
American Indian or Alaska Native	0 (0)
Asian	1 (5)
Black or African American	0 (0)
Hispanic/Latino	4 (20)
Native Hawaiian or other Pacific Islander	1 (5)
White	14 (70)
Household composition, n (%)	
I live alone	7 (35)
I live with my partner - (currently) no children in the house	3 (15)
I live with my children - (currently) no partner in the house	1 (5)
I live with my partner and child(ren)	6 (30)
I live with a family member(s)	3 (15)
I live with a roommate(s)	0 (0)
Other	0 (0)
Employment, n (%)	
Working full-time	14 (70)
Working part-time	2 (10)
Homemaker	2 (10)
Full-time student	1 (5)
Retired	1 (5)
Education, n (%)	
High school degree or equivalent (eg, GED[a])	1 (5)
Some college, no degree	5 (25)
Associate degree	1 (5)
Bachelor’s degree	8 (40)
Master’s degree	3 (15)
Doctoral degree	2 (10)
Annual household income (US$), n (%)	
<$25,000	0 (0)
$25,000-$49,999	3 (15)
$50,000-$74,999	4 (20)
$75,000-$99,999	3 (15)
$100,000-$124,999	2 (10)
$125,000-$149,999	2 (10)

$150,000-$174,999	0 (0)
$175,000-$199,999	0 (0)
≥$200,000	5 (25)
Would rather not answer this question	1 (5)

[a]GED: General Educational Development.

Objective Sleep

Nightly objective measurement of sleep (n=676 nights nested within 20 participants) revealed multiple improvements when participants used the goggles at bedtime. Key findings included increased sleep duration (+12 min, *P*=.01); increased deep sleep, reflected both in duration (+6 min, *P*=.002) and proportion of the night (7% relative increase, *P*=.02); enhanced BodyScore (+4%, *P*=.002); fewer nighttime awakenings (−7%, *P*=.02); reduced total wake time at night after sleep onset (−6 min, *P*=.047); and improved SleepScore, indicating overall sleep quality (+3%, *P*=.02). Detailed objective sleep metrics are displayed in Table 2.

Table . Multilevel regression results for objective sleep (n=676 nights nested within 20 participants), comparing the baseline period (sleep at home prior to intervention) to the intervention period (sleep at home using smart goggles at bedtime).

Outcomes	Objective measurement of sleep, observed mean (SD)[a]		Estimated marginal means[b]		
	Baseline period	Intervetion period	Intercept (SE)	β[c]	*P* value
Total sleep time in minutes	331.55 (57.49)	343.48 (62.69)	331.52 (4.87)	12.04	.01
Sleep onset latency in minutes	23.34 (18.33)	23.20 (18.17)	23.28 (1.64)	−0.13	.92
Number of awakenings	5.51 (2.10)	5.14 (2.12)	5.51 (0.17)	−0.38	.02
Time spent awake after sleep onset in minutes	60.50 (31.30)	54.32 (31.38)	60.17 (2.72)	−5.39	.047
Light sleep in minutes	215.36 (49.64)	219.86 (50.57)	215.47 (3.99)	4.54	.26
Deep sleep in minutes	57.52 (19.92)	63.19 (24.27)	57.66 (1.79)	5.58	.002
REM[d] sleep in minutes	58.67 (23.54)	60.44 (25.55)	58.52 (1.93)	1.94	.32
Percentage of time spent awake after sleep onset	16 (7)	14 (7)	15.36 (0.58)	−1.02	.08
Percentage of time in light sleep	54 (8)	55 (8)	54.50 (0.60)	0.02	.97
Percentage of time in deep sleep	15 (6)	16 (7)	15.10 (0.49)	1.14	.02
Percentage of time in REM sleep	15 (6)	15 (6)	14.92 (0.46)	−0.12	.80
SleepScore[e]	69.57 (10.89)	71.65 (10.87)	69.61 (0.87)	2.02	.02
BodyScore[e]	73.14 (10.38)	76.13 (11.15)	73.42 (0.86)	2.64	.002
MindScore[e]	69.30 (13.76)	69.68 (14.48)	69.18 (1.13)	0.55	.62

[a]For the baseline and intervention periods, each mean was calculated by averaging nights across participants, then averaging those participants' averages to a single simple average.

[b]These are the outcomes of separate multilevel regression analyses. Each row shows results from a different single-predictor, single-outcome model.

[c]The beta values are unstandardized and can therefore be interpreted on the same scale as the original data.

[d]REM: rapid eye movement.

[e]These scores range from 0 to 100. SleepScore is an age- and gender-normalized measure of overall sleep quality, BodyScore is an age- and gender-normalized measure of deep sleep, and MindScore is an age- and gender-normalized measure of REM sleep.

Self-Reported Sleep Quality and Well-Restedness

Multilevel regression analyses of daily self-reported sleep data (N=723 nights nested within 20 participants) showed that participants perceived better sleep quality (β=12.37, *P*<.001) and felt more rested in the morning (β=12.13, *P*<.001) when

using the goggles at bedtime compared to baseline, as assessed on a scale from 0 to 100.

Sleepiness, Anxiety, Stress, and Relaxation

Across 334 nights of the intervention period, multilevel regression analyses comparing responses immediately after using the goggles to those reported immediately before showed that participants felt sleepier (β=9.98, $P<.001$), less stressed (β=−10.38, $P<.001$), less anxious (β=−12.87, $P<.001$), and more relaxed (β=11.76, $P<.001$), all rated on a scale from 0 to 100.

At the end of the intervention period, compared to the end of the baseline period, participants' scores on the 6-item STAI showed reduced anxiety ($t_{16}=2.31$, $P=.03$), reflecting a 10% decrease and confirming the nightly analyses.

Discussion

Nonpharmacological techniques for promoting relaxation and improving sleep have the potential to help millions of individuals who experience suboptimal sleep [20]. This study evaluated the effectiveness of smart goggles designed to induce relaxation and support healthy sleep. Outcomes were measured using both self-reported data and a polysomnography-validated, noncontact biomotion device. The study population included a nonclinical sample of adults reporting poor sleep in the absence of diagnosed sleep disorders.

Compared to baseline, using the goggles at bedtime led to objective improvements in both sleep quality and duration. Although total sleep time remained less than 6 hours per night on average, objective improvements were seen in several parameters, including:

- sleep duration;
- increased deep sleep, both in duration and as a proportion of the night;
- reduced number of nighttime awakenings;
- decreased time spent awake at night after initially falling asleep; and
- enhanced sleep quality.

Aligning with these objective results, questionnaire data showed that participants perceived improvement in their sleep quality and felt more well-rested in the morning. In addition, immediately after using the goggles, participants felt sleepier, less stressed, less anxious, and more relaxed, compared to their experience immediately before using the goggles. A standardized inventory administered before and after the 3-week intervention period also indicated reduced anxiety, confirming the nightly analysis.

The observed improvements in the objective and self-reported sleep data may be attributed to increased relaxation resulting from the use of the goggles at bedtime. Vibration has been previously shown to be able to induce physiological relaxation [9,10] and support sleep [11-13]. This interpretation is supported by the changes in perceived relaxation, stress, and anxiety. However, objective parameters of physiological relaxation prior to sleep were not assessed, presenting an interesting avenue for future research.

Further studies could explore the intervention through a controlled trial, including comparison to goggles without vibration or to other relaxation techniques. This study assessed changes in parameters from the baseline to the intervention period while using the product at home, resembling how it is used outside of a research setting. The within-participants study design, which included long-term product use following a baseline period without the intervention, provides confidence that significant effects are due to the intervention itself. Although this study design has limitations, it reflects the real-world experience of introducing an intervention into the home environment, while also accounting for night-to-night variations in sleep patterns.

While longitudinal data provides an advantage by enabling the detection of nightly within-person differences, this pilot study is limited by its small sample size. Although we compared participants' improvements in sleep and related outcomes to their own baseline, the small sample size could have introduced a potential bias, and therefore these results should be interpreted with caution. Future research should investigate the effects of this intervention using larger sample sizes. Another limitation is that contactless technology is not ambulatory and therefore does not capture objective measures of daytime activity. However, the noncontact sleep measurement system used in this study offers the advantage of significantly higher overall accuracy, particularly in specificity, compared to traditional research-based actigraphy [16].

To conclude, this study found that using vibration therapy administered via smart goggles before going to sleep was associated with improvements in both objective sleep measures and self-reported outcomes. While these findings must be interpreted with caution, our data suggest that a device delivering gentle vibrations to the eye and temple area may have the potential to promote relaxation, decrease anxiety, and support healthy sleep when used at bedtime by adults with suboptimal sleep. Not only did sleep quality and sleep duration increase relative to baseline, but there was a decrease in the number of awakenings and the duration of time spent awake during the night. Further, improvement was seen in multiple metrics of deep sleep, which is vital for brain health and physical recovery. Future research expanding on this pilot study is warranted to confirm the preliminary evidence presented in this study.

Acknowledgments

Therabody provided the SmartGoggles and funding support to SleepScore Labs to design, conduct, and report the study.

Data Availability

The data analyzed in this study may be available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors are employed by SleepScore Labs.

References

1. Grandner MA. Epidemiology of insufficient sleep and poor sleep quality. In: Grandner MA, editor. Sleep and Health: Sleep Health Acad Press; 2019:11-20. [doi: 10.1016/B978-0-12-815373-4.00002-2]
2. Consensus Conference Panel, Watson NF, Badr MS, et al. Recommended amount of sleep for a healthy adult: a joint consensus statement of the American Academy of Sleep Medicine and Sleep Research Society. J Clin Sleep Med 2015 Jun 15;11(6):591-592. [doi: 10.5664/jcsm.4758] [Medline: 25979105]
3. Chellappa SL, Aeschbach D. Sleep and anxiety: from mechanisms to interventions. Sleep Med Rev 2022 Feb;61:101583. [doi: 10.1016/j.smrv.2021.101583] [Medline: 34979437]
4. Guillodo E, Lemey C, Simonnet M, et al. Clinical applications of mobile health wearable-based sleep monitoring: systematic review. JMIR Mhealth Uhealth 2020 Apr 1;8(4):e10733. [doi: 10.2196/10733] [Medline: 32234707]
5. Ko PRT, Kientz JA, Choe EK, Kay M, Landis CA, Watson NF. Consumer sleep technologies: a review of the landscape. J Clin Sleep Med 2015 Dec 15;11(12):1455-1461. [doi: 10.5664/jcsm.5288] [Medline: 26156958]
6. Watson NF, Lawlor C, Raymann RJEM. Will consumer sleep technologies change the way we practice sleep medicine? J Clin Sleep Med 2019 Jan 15;15(1):159-161. [doi: 10.5664/jcsm.7596] [Medline: 30621844]
7. Bertisch SM, Wells RE, Smith MT, McCarthy EP. Use of relaxation techniques and complementary and alternative medicine by American adults with insomnia symptoms: results from a national survey. J Clin Sleep Med 2012 Dec 15;8(6):681-691. [doi: 10.5664/jcsm.2264] [Medline: 23243402]
8. Maust DT, Solway E, Clark SJ, Kirch M, Singer DC, Malani P. Prescription and nonprescription sleep product use among older adults in the United States. Am J Geriatr Psychiatry 2019 Jan;27(1):32-41. [doi: 10.1016/j.jagp.2018.09.004] [Medline: 30409547]
9. Zhou Y, Murata A, Watanabe J. The calming effect of heartbeat vibration. Presented at: 2020 IEEE HAPTICS Symposium (HAPTICS); Mar 28-31, 2020; Crystal City, VA. [doi: 10.1109/HAPTICS45997.2020.ras.HAP20.157.5a2e1551]
10. Zhang N, Fard M, Bhuiyan MHU, Verhagen D, Azari MF, Robinson SR. The effects of physical vibration on heart rate variability as a measure of drowsiness. Ergonomics 2018 Sep;61(9):1259-1272. [doi: 10.1080/00140139.2018.1482373] [Medline: 29871584]
11. Himes BJ, Blotter JD, Kay DB, et al. The effect of beat frequency vibration on sleep latency and neural complexity: a pilot study. IEEE Trans Neural Syst Rehabil Eng 2021;29:872-883. [doi: 10.1109/TNSRE.2021.3076983] [Medline: 33950842]
12. Choi SH, Yoon H, Jin HW, et al. Effect of closed-loop vibration stimulation on heart rhythm during naps. Sensors (Basel) 2019 Sep 24;19(19):4136. [doi: 10.3390/s19194136] [Medline: 31554268]
13. Choi SH, Kwon HB, Jin HW, et al. Weak closed-loop vibrational stimulation improves the depth of slow-wave sleep and declarative memory consolidation. Sleep 2021 Jun 11;44(6):zsaa285. [doi: 10.1093/sleep/zsaa285] [Medline: 33367712]
14. O'Hare E, Flanagan D, Penzel T, Garcia C, Frohberg D, Heneghan C. A comparison of radio-frequency biomotion sensors and actigraphy versus polysomnography for the assessment of sleep in normal subjects. Sleep Breath 2015 Mar;19(1):91-98. [doi: 10.1007/s11325-014-0967-z] [Medline: 24614968]
15. Schade MM, Bauer CE, Murray BR, et al. Sleep validity of a non-contact bedside movement and respiration-sensing device. J Clin Sleep Med 2019 Jul 15;15(7):1051-1061. [doi: 10.5664/jcsm.7892] [Medline: 31383243]
16. Chinoy ED, Cuellar JA, Huwa KE, et al. Performance of seven consumer sleep-tracking devices compared with polysomnography. Sleep 2021 May 14;44(5):zsaa291. [doi: 10.1093/sleep/zsaa291] [Medline: 33378539]
17. Ohayon MM, Carskadon MA, Guilleminault C, Vitiello MV. Meta-analysis of quantitative sleep parameters from childhood to old age in healthy individuals: developing normative sleep values across the human lifespan. Sleep 2004;27:1255-1273. [doi: 10.1093/sleep.27.7.1255] [Medline: 15586779]
18. Lesage FX, Berjot S, Deschamps F. Clinical stress assessment using a visual analogue scale. Occup Med (Lond) 2012 Dec;62(8):600-605. [doi: 10.1093/occmed/kqs140] [Medline: 22965867]
19. Marteau TM, Bekker H. The development of a six-item short-form of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI). Br J Clin Psychol 1992 Sep;31(3):301-306. [doi: 10.1111/j.2044-8260.1992.tb00997.x] [Medline: 1393159]
20. Maness DL, Khan M. Nonpharmacologic management of chronic insomnia. Am Fam Physician 2015 Dec 15;92(12):1058-1064. [Medline: 26760592]

Abbreviations

NREM-3: non–rapid eye movement stage 3

REM: rapid eye movement
STAI: State-Trait Anxiety Inventory

Edited by A Mavragani; submitted 15.03.24; peer-reviewed by GR Elkins, R Chamorro; revised version received 21.10.24; accepted 29.10.24; published 03.01.25.

Please cite as:
Danoff-Burg S, Gottlieb E, Weaver MA, Carmon KC, Lara Ledesma D, Rus HM
Effects of Smart Goggles Used at Bedtime on Objectively Measured Sleep and Self-Reported Anxiety, Stress, and Relaxation: Pre-Post Pilot Study
JMIR Form Res 2025;9:e58461
URL: https://formative.jmir.org/2025/1/e58461
doi:10.2196/58461

The Use of Social Media on Enhancing Dental Care and Practice Among Dental Professionals: Cross-Sectional Survey Study

Joseph Macadaeg Acosta[1,2], MDE, DMD; Palinee Detsomboonrat[3], MSc, DDS, PhD; Pagaporn Pantuwadee Pisarnturakit[3], MSc, DDS, PhD; Nipaporn Urwannachotima[3], MEd, DDS, PhD

[1]International Graduate Program in Dental Public Health, Department of Community Dentistry, Faculty of Dentistry, Chulalongkorn University, Bangkok, Thailand

[2]School of Dentistry, University of Baguio, Baguio City, Philippines

[3]Faculty of Dentistry, Department of Community Dentistry, Chulalongkorn University, 34 Henri Dunant Road, Pathumwan, Bangkok, Thailand

Corresponding Author:
Nipaporn Urwannachotima, MEd, DDS, PhD
Faculty of Dentistry, Department of Community Dentistry, Chulalongkorn University, 34 Henri Dunant Road, Pathumwan, Bangkok, Thailand

Abstract

Background: As digitalization continues to advance globally, the health care sector, including dental practice, increasingly recognizes social media as a vital tool for health care promotion, patient recruitment, marketing, and communication strategies.

Objective: This study aimed to investigate the use of social media and assess its impact on enhancing dental care and practice among dental professionals in the Philippines.

Methods: A cross-sectional survey was conducted among dental practitioners in the Philippines. The study used a 23-item questionnaire, which included 5 questions on dentists' background and demographic information and 18 questions regarding the use, frequency, and purpose of social media in patient advising and quality of care improvement. Data were analyzed using SPSS software, with frequency distributions and χ^2 tests used to assess the association between social media use and demographic variables and the impact on dental practice.

Results: The 265 dental practitioners in this study were predominantly female (n=204, 77%) and aged between 20 - 30 years (n=145, 54.7%). Most of the participants were general practitioners (n=260, 98.1%) working in a private practice (n=240, 90.6%), with 58.5% (n=155) having 0 - 5 years of clinical experience. Social media use was significantly higher among younger practitioners (20 - 30 years old) compared to older age groups (P<.001), though factors such as sex, dental specialty, and years of clinical practice did not significantly influence use. The majority (n=179, 67.5%) reported using social media in their practice, primarily for oral health promotion and education (n=191, 72.1%), connecting with patients and colleagues (n=165, 62.3%), and marketing (n=150, 56.6%). Facebook (n=179, 67.5%) and YouTube (n=163, 61.5%) were the most frequented platforms for clinical information, with Twitter (subsequently rebranded X) being the least used (n=4, 1.5%). Despite widespread social media engagement, only 8.7% (n=23) trusted the credibility of web-based information, and 63.4% (n=168) perceived a potential impact on the patient-dentist relationship due to patients seeking information on the internet. Social media was also perceived to enhance practice quality, with users reporting significant improvements in patient care (P=.001).

Conclusions: The findings highlight that social media is widely used among younger dental practitioners, primarily for education, communication, and marketing purposes. While social media use is associated with perceived improvements in practice quality and patient care, trust in information on social media remains low, and concerns remain regarding its effect on patient relationships. It is recommended to establish enhanced guidelines and provide reliable web-based resources to help dental practitioners use social media effectively and responsibly.

(JMIR Form Res 2025;9:e66121) doi:10.2196/66121

KEYWORDS

social media; oral health promotion; oral health education; dentists; dental practice; dental professionals; dental practitioners

Introduction

Traditionally, dental practices advertised their services primarily through local yellow pages, such as telephone directories listing local businesses or by placing promotional notices in the windows of dental offices [1]. Word-of-mouth was also a common method, where information about the quality of service was shared from one person to another [2]. However, with the advent of technological advancements, there has been a significant shift toward the use of social media. Social media

platforms offer the potential to reach diverse audiences, including those seeking information on oral health management, thereby promoting oral health [3].

Social media refers to a group of web-based applications built on the foundations of Web 2.0 [4,5]. According to the Merriam-Webster Dictionary, social media is defined as "any form of electronic communication through which users create web-based communities to share information, personal messages, ideas, and other content such as photos and videos" [6]. Popular platforms include social networking sites like Facebook and Twitter (subsequently rebranded X), media-sharing sites like YouTube and Instagram, blogs, and microblogging sites [7]. These platforms are widely used for searching, sharing, and communicating information, including health-related content [8,9]. Facebook fosters connections through its friend-based network, creating a web-based community where users can stay in touch and share content [10]. Twitter's microblogging feature allows users to share ideas and opinions in real time [11,12]. YouTube is a leading video-sharing platform [13-15], while Instagram, a photo and video social networking site, facilitates real-time communication through comments and direct messages and supports professional engagement and knowledge sharing [16-18]. TikTok, known for its short video content, has also gained popularity [19,20].

Social media serves as a dynamic tool for exchanging information and communicating with colleagues, patients, and the public on health-related issues. Public health professionals and organizations use these platforms for a variety of purposes, including health education, telemedicine, social marketing, scientific research, recruitment, career development, and professional networking [9]. Dental practitioners are increasingly leveraging social media to educate and inform the public by sharing detailed content on various dental procedures and educational activities [8,21-25], providing oral hygiene instructions and promoting oral health [26,27], enhancing communication [28-30], and engaging in marketing efforts [1,25,31,32]. Given the widespread use of social media, using these platforms responsibly and effectively in health care is crucial to ensure that they contribute positively to patient health outcomes.

As of April 2024, there were 5.44 billion internet users worldwide, representing 67.1% of the global population. Of this total, 5.07 billion, or 62.6% of the world's population, were social media users [33]. Based on the number of monthly active users, the most popular social networks globally are Facebook, YouTube, Instagram, WhatsApp, and TikTok [34]. In the Philippines, approximately 68.72 million people are social media users, with an average daily use of over 3 hours, the highest across the Asia-Pacific region. Facebook, Instagram, TikTok, and X (formerly known as Twitter) are currently some of the most popular platforms, providing opportunities for connecting with family and friends, sharing content via digital platforms, and expanding the reach of promotional marketing through web-based banners [35]. Social media in the Philippines has gained significant traction for entertainment, communication, marketing, and professional education. In dentistry, practitioners use these platforms to share content related to oral health, modern treatment options, trends in dental materials, and treatment costs.

Although several reports have highlighted the use of social media among dentists and its influence on dental practice [3,8,9,22,30,32,36,37], as well as its impact on dental education among faculty and students [7,21,23,38,39], relatively few studies have examined the use and influence of social media on dental practice in Southeast Asia, particularly in the Philippines where these platforms are widely used. This study aims to investigate the use of social media and assess its impact on enhancing dental care and practice among dental professionals in the Philippines.

Methods

Design and Study Population

This study used a cross-sectional survey design with purposive sampling, targeting all eligible licensed dental professionals, including both general dentists and specialists working in private or public clinical settings and who were involved in patient care, education, and practice management. Participants were also required to be active members of the Philippine Dental Association-Baguio City Chapter for the 2023 - 2024 term. The study population consisted of 265 dental practitioners. Practitioners who were not active members of the chapter were excluded from the study.

The survey questionnaire was adapted from previously published literature related to social media use [22,36,38]. For this study, social media included popular platforms such as Facebook, YouTube, Twitter, Instagram, TikTok, and Viber, which dental practitioners may have used for professional purposes like patient education, practice marketing, peer communication, and access to clinical information. A panel of experts validated the questionnaire items and the questionnaire underwent pretesting to ensure applicability and reliability. Revisions and minor adjustments were made based on the pilot study results and expert feedback. The 23-item questionnaire was divided into two sections. The first section gathered background and demographic information, including age, sex, dental specialty, work sector, and years of clinical practice. The second section comprised 18 questions designed to gather information on social media use; frequency of use; and purpose in dental practice, particularly its role in advising patients and improving the quality of care. This study was conducted within dental practice settings, with participants recruited through web-based methods. Recruitment emphasized voluntary participation, with all respondents providing informed consent before inclusion in the study. The questionnaire was administered via Google Forms, with the survey link disseminated through email, a Facebook page, and the Baguio City Chapter group chat. Respondents were encouraged to provide suggestions, assist in recruitment, and further share the questionnaire with other dental practitioners to reach more participants. They were given one day to complete the form.

Data Analysis

Data were analyzed using SPSS software version 28.0 (IBM Corp). The analysis included frequency distributions that were

expressed in numbers and percentages, as well as χ^2 tests to assess the association between social media use and demographic or social variables and the impact on dental practice. Statistical significance was set at an α level of .05. This study adhered to the reporting guidelines for cross-sectional studies following the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement.

Ethical Considerations

Ethical approval was obtained from the Human Research Ethics Committee of the Faculty of Dentistry, Chulalongkorn University (Study Code: HREC-DCU 2023 - 108). Informed consent was obtained from all participants through a written consent form, which provided details about the study's objectives and potential benefits. Participation in the study was entirely voluntary, with participants having the right to withdraw or discontinue at any time without any penalties or loss of benefits. All information was kept strictly confidential and anonymized to ensure participants could not be identified. No financial rewards or incentives were provided for participation.

Results

The background and demographic details of the 265 study participants are summarized in Table 1. The majority of participants were female (n=204, 77%), with over half (n=145, 54.7%) aged between 20 to 30 years. This was followed by participants aged 31 - 40 years (n=52, 19.6%), and both groups of participants aged 41 - 50 years and over 50 years represented 12.8% (n=34) of participants. Most participants (n=260, 98.1%) were general practitioners, and 90.6% (n=240) worked as private dentists. Regarding clinical experience, 58.5% (n=155) had 0 to 5 years of practice, while 29.8% (n=79) had over 10 years, and 11.7% (n=31) had 6 to 10 years of practice.

Table . Demographic variables and their association with the use of social media among dental practitioners.

Variable	Use of social media, n (%)[a]			Total (n=265), n (%)	*P* value[b]
	Yes	Sometimes	No		
Age (years)					<.001
20 - 30	102 (70.3)	35 (24.1)	8 (5.5)	145 (54.7)	
31 - 40	31 (60)	21 (40)	0 (0)	52 (19.6)	
41 - 50	27 (79)	2 (6)	5 (15)	34 (12.8)	
>50	19 (56)	15 (44)	0 (0)	34 (12.8)	
Sex					.66
Male	40 (66)	19 (31)	2 (3)	61 (23)	
Female	139 (68.1)	54 (26.5)	11 (5.4)	204 (77)	
Dental specialty					.25
General practitioners	177 (68.1)	70 (26.9)	13 (5)	260 (98.1)	
Specialists	2 (40)	3 (60)	0 (0)	5 (1.9)	
Working sector					.30
Government	10 (71)	2 (14)	2 (14)	14 (5.3)	
Private	160 (66.6)	69 (28.8)	11 (4.6)	240 (90.6)	
Both (government and private)	9 (82)	2 (18)	0 (0)	11 (4.2)	
Number of years of clinical practice					.19
0 to 5	109 (70.3)	38 (24.5)	8 (5.2)	155 (58.5)	
6 to 10	24 (77)	7 (23)	0 (0)	31 (11.7)	
>10	46 (58)	28 (35)	5 (6)	79 (29.8)	

[a]Percentages were calculated based on the number of participants corresponding to each variable.

[b]Chi-square test; α=.05.

There was a significant increase in social media use among the younger age group (20 - 30 years) compared to older age groups (31 - 40 years, 41 - 50 years, and >50 years), with a *P* value of <.001. However, sex, dental specialty, work sector, and years of clinical practice did not significantly influence social media use.

Table 2 illustrates the frequency distribution of social media use among dental practitioners. Of the 265 participants, most reported moderate use of social media with 37.4% (n=99) using it 1 - 2 days per week. However, nearly two-thirds (n=164, 61.9%) reported using social media at least 3 days weekly. Facebook (n=179, 67.5%) and YouTube (n=163, 61.5%) were the most commonly visited platforms for obtaining clinical information, followed by Instagram (n=115, 43.4%). Twitter

(n=4, 1.5%) was the least popular platform. Social media was mainly used for oral health education and promotion (n=191, 72.1%); communication with patients, friends, and family (n=165, 62.3%); and marketing and advertising (n=150, 56.6%).

Table . Frequency, preferred platforms, and purpose of social media use.

Variables	Frequency (n=265), n (%)
Social media use	
Never	2 (0.8)
Seldom (average of 1 - 2 d/wk)	99 (37.4)
Occasionally (average of 3 - 5 d/wk)	96 (36.2)
Frequently (average of 6 - 7 d/wk)	68 (25.7)
Social media website used[a]	
Twitter	4 (1.5)
Facebook	179 (67.5)
Instagram	115 (43.4)
YouTube	163 (61.5)
TikTok	38 (14.3)
Viber	23 (8.7)
Others	75 (28.3)
Purpose of social media use[a]	
Oral health promotion and education	191 (72.1)
Connect and communicate with patients, friends, and family	165 (62.3)
Exchange opinions and views regarding cases with colleagues	113 (42.6)
Marketing and advertising	150 (56.6)

[a]Given that study participants could indicate multiple responses, the sum of the percentages is not 100.

The responses of dental practitioners to questions regarding social media use are presented in Figure 1. Of the 265 participants, 67.5% (n=179) indicated that they use social media in their dental practice. Despite this, 83% (n=220) of dentists did not have their own web-based practice forums or websites. However, the majority (n=227, 85.7%) reported using social media to communicate with other dental professionals, and 41.5% (n=110) used it for marketing purposes. Although 41.1% (n=109) of dentists referred to articles or research information from social media for clinical practice, nearly half (n=127, 47.9%) were reluctant to broadcast treatment outcomes on the internet to attract patients, and 41.9% (n=111) were open to providing web-based consultations. Only 8.7% (n=23) of dentists trusted the credibility of information on social media, while 63.4% (n=168) believed that social media can affect the patient-dentist relationship when patients seek information on the internet. Additionally, most participants (n=154, 58.1%) did not allow patients to access their information through a website, but 64.2% (n=170) would recommend a trusted website to patients.

Table 3 explores the relationship between social media use and its impact on dental practice. Participants who used social media in their practice believed that it enhanced the quality of care provided to patients (*P*=.001) and significantly reported that it improved their dental practice (*P*=.02).

Figure 1. Percentage of dental practitioners' responses to questions on social media use (n=265).

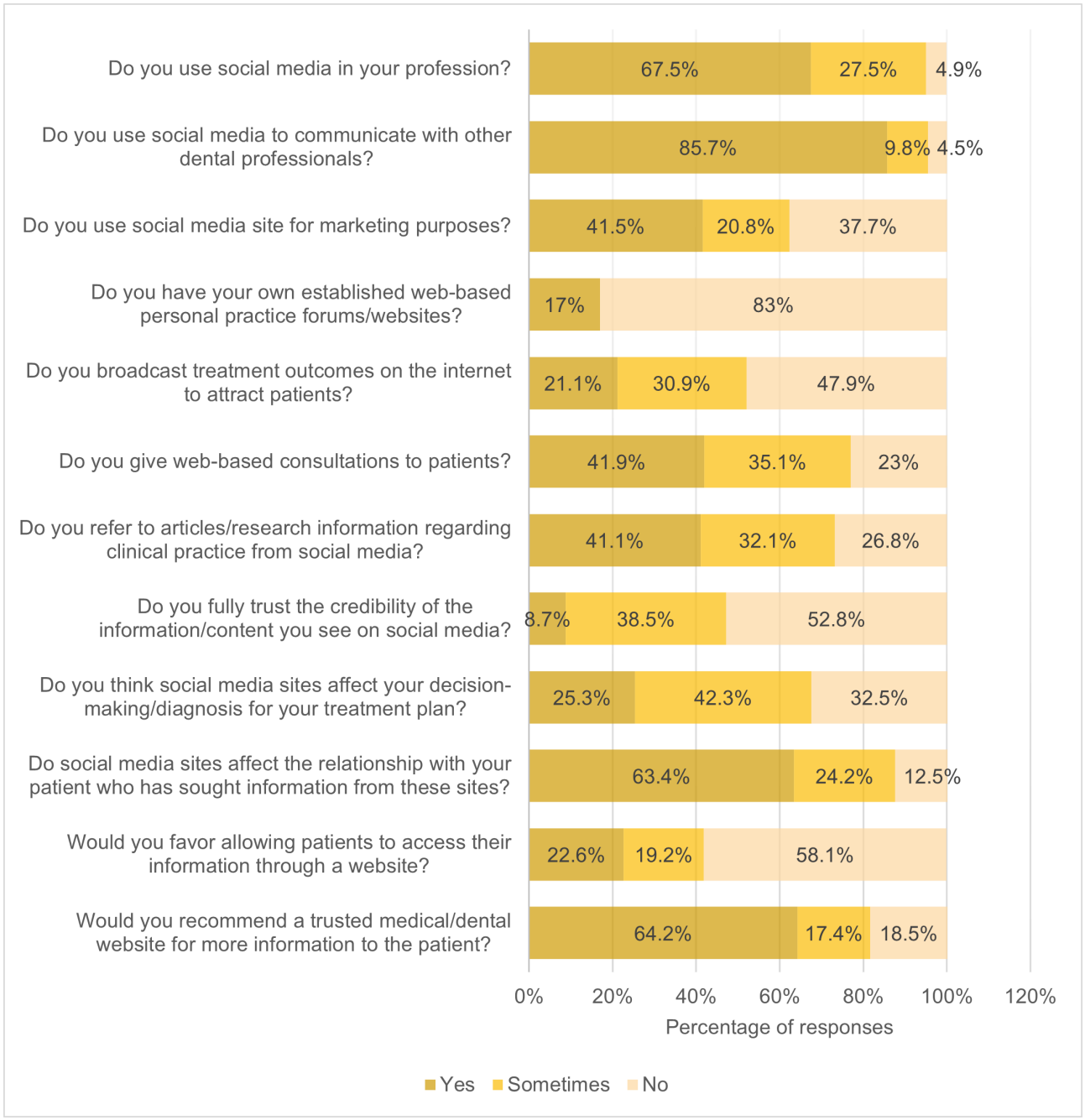


Table . Use of social media and respondents' perceived impact on enhancing patient care quality and dental practice.

Variable	Use of social media, n (%)[a]			Total (n=265), n (%)	*P* value[b]
	Yes	Sometimes	No		
Social media improves the quality of care delivered to patients					.001
Yes	126 (74.1)	40 (23.5)	4 (2.4)	170 (64.2)	
Maybe	42 (56)	24 (32)	9 (12)	75 (28.3)	
No	11 (55)	9 (45)	0 (0)	20 (7.5)	
Social media improves dental practice					.02
Yes	106 (74.7)	30 (21.1)	6 (4.2)	142 (53.6)	
Maybe	48 (57)	29 (35)	7 (8)	84 (31.7)	
No	25 (64)	14 (36)	0 (0)	39 (14.7)	

[a]Percentages were calculated based on the number of participants corresponding to each variable.

[b]Chi-square test was used with an α level of .05.

Discussion

Principal Findings

This study aimed to investigate the use of social media and assess its impact on enhancing dental care and practice among dental professionals. The findings revealed that over half of the participants aged 20 - 30 years are actively using social media in their dental profession, aligning with other studies that indicate younger dentists or health care professionals are more likely to use social media compared to their older counterparts [36,38,40-42]. Variables such as sex, dental specialty, work sector, and years of clinical practice did not significantly influence social media use.

Most respondents reported using social media 1 - 2 days per week to engage with the public and other dental professionals. This is in contrast to other studies where social media was used at least once daily for obtaining information [32,42]. Among social media users in the Philippines, Facebook and YouTube were identified as the primary platforms for accessing clinical information, which is consistent with The Digital Report 2024, highlighting that Facebook has 86.75 million active users in the Philippines [43]. These findings are similar to those reported among dentists in the United Kingdom and the United States, where Facebook is the most commonly used platform in dental practice [1,31]. This was also true as among dental students [7]. In comparison, dentists in Saudi Arabia predominantly use Twitter [22,25] and Instagram [32], while WhatsApp is the preferred platform among Indian dentists [37]. Similarly, South African dentists primarily use Google Plus and Facebook for personal rather than professional purposes [42].

Previous studies have shown that marketing is a common reason for using social media [1,25,31,32]. In our study, 72.1% (191/265) of participants indicated that their primary reason for using social media was for oral health promotion and education. This finding is consistent with earlier research, which found that a high proportion of respondents use social media for dental education and prefer to share clinical work [21]. Other studies also support the notion that education and sourcing oral health information are primary purposes for social media use among dental professionals [22-25]. Social media interventions have positively impacted various aspects of oral health, particularly among adolescents, where platforms like YouTube, WhatsApp, Instagram, and Telegram have been effectively used to promote oral health [27]. For instance, a study demonstrated that oral health education via Telegram significantly improved oral health outcomes among secondary school students [44] and adolescents [45] in Iran. Additionally, Instagram has been shown to support orthodontists in motivating young patients to maintain proper oral hygiene [46], and WhatsApp has been integrated successfully into oral hygiene protocols to enhance patient compliance and oral health during orthodontic treatment [47]. Presenting audiovisual information through YouTube has also significantly improved patient knowledge [48].

Another key purpose of social media use is communication and the exchange of opinions with colleagues. According to the survey, 62.3% (165/265) of participants used social media to connect with patients, friends, and family, while 85.7% (227/265) used it to communicate with other dental professionals. This is supported by studies showing that social media communication, such as Instagram, can reduce patient anxiety prior to dental procedures [28], enhance dentists' communication skills with other professionals [38], and improve patient interactions [30]. Another study found that a significant proportion of the younger population engages with social media to discuss dentistry. This highlights the need for careful management to ensure the dissemination of accurate dental health information [25].

The respondents of the survey also indicated that social media had the potential to enhance dental practice and the quality of care provided to patients. This finding is in agreement with study that shows that dentists who use social media in their practice not only recommend it to their peers but also believe it enhances their practice [22]. Social media can improve various aspects of dental practice, including service provision, advertising, counseling, and oral health education, and it can also be a tool for professional development [49]. Dental organizations and educators can use social media to disseminate information and updates [36,50]. Furthermore, multidimensional health care approaches, including social media, have been

proven highly successful in improving patient care, increasing public knowledge, facilitating research, connecting health care providers, improving medical education, and addressing public health crises [29].

This study has several limitations that should be considered when interpreting the results. First, the study population was limited to dental practitioners in Baguio City, Philippines, which may restrict the generalizability of the findings to the broader dentist population in the country. The uneven distribution of survey participation across various demographic variables, such as age, sex, dental specialty, working sector, and years of clinical practice, may further impact the ability to extrapolate the results to the entire dental community. Additionally, the use of an electronic survey as the sole data collection method presented inherent challenges. A significant portion of the target population may not respond to or use electronic surveys, potentially introducing nonresponse bias and affecting the representativeness of the sample. This bias could be particularly relevant in the context of social media use, as those who are less inclined to participate in electronic surveys may also be less likely to engage with social media platforms. Furthermore, the cross-sectional nature of the study limits the ability to establish causal relationships between social media use and its impact on dental practice and patient care. Longitudinal studies would be necessary to better understand the long-term effects of social media integration in dental settings. Despite these limitations, this study provides valuable insights into the current use of social media among dental practitioners in the Philippines and its perceived impact on dental care and practice. Future research should aim to include larger, more representative samples across different regions and settings to enhance the generalizability of the findings. Additionally, mixed-methods approaches combining quantitative and qualitative data collection techniques could provide a more comprehensive understanding of the topic.

Conclusions

Social media is widely used among younger dental practitioners, primarily for education, communication, and marketing, with a notable impact on enhancing practice quality and patient care. Social media platforms, especially Facebook and YouTube, were commonly used by practitioners to promote oral health, engage with patients and colleagues, and support marketing efforts.

Social media offers a convenient space for professional growth, patient education, and community interaction, which many dentists perceive as beneficial to their practice. However, a low trust in social media information and concerns about its influence on patient-dentist relationships indicate the need for clear guidelines and quality web-based resources. Enhanced support in these areas can help dental professionals maximize the positive impacts of social media on dental care and practice.

Future studies should compare social media with traditional marketing methods in the dental practice such as print ads, community outreach, and patient referrals. These traditional methods remain valuable for building trust, local credibility, and reaching offline patients. Comparing these strategies can help find the best way to combine digital and traditional approaches for better patient engagement and practice success.

Acknowledgments

The authors would like to thank all participants who participated in this study. We also appreciated the funding support from the Faculty of Dentistry at Chulalongkorn University.

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

References

1. Parmar N, Dong L, Eisingerich AB. Connecting with your dentist on Facebook: patients' and dentists' attitudes towards social media usage in dentistry. J Med Internet Res 2018 Jun 29;20(6):e10109. [doi: 10.2196/10109] [Medline: 29959108]
2. Jung YS, Yang HY, Choi YH, et al. Factors affecting use of word-of-mouth by dental patients. Int Dent J 2018 Oct;68(5):314-319. [doi: 10.1111/idj.12387] [Medline: 29572927]
3. Mary AV, Kesavan R, V V, Pi J, G D, S S. Impact of social media on dental treatment needs: a cross sectional pilot study. Int J Appl Dent Sci 2023 Jan 1;9(4):313-317. [doi: 10.22271/oral.2023.v9.i4e.1887]
4. Afful-Dadzie E, Afful-Dadzie A, Egala SB. Social media in health communication: a literature review of information quality. Hlth Inf Manag 2023 Jan;52(1):3-17. [doi: 10.1177/1833358321992683] [Medline: 33818176]
5. Kaplan AM, Haenlein M. Users of the world, unite! The challenges and opportunities of social media. Bus Horiz 2010 Jan;53(1):59-68. [doi: 10.1016/j.bushor.2009.09.003]
6. Social media. Merriam-Webster Dictionary. URL: https://www.merriam-webster.com/dictionary/social%20media [accessed 2024-12-17]

7. Kenny P, Johnson IG. Social media use, attitudes, behaviours and perceptions of online professionalism amongst dental students. Br Dent J 2016 Nov 18;221(10):651-655. [doi: 10.1038/sj.bdj.2016.864] [Medline: 27857111]
8. Sivaramakrishnan G, AbdulAmeer F, Faisal F, et al. Use of social media to view and post dentistry-related information in Bahrain: a cross-sectional study. Healthc Inform Res 2023 Jan;29(1):31-39. [doi: 10.4258/hir.2023.29.1.31] [Medline: 36792099]
9. Freire Y, Sánchez MG, Suárez A, Joves G, Nowak M, Díaz-Flores García V. Influence of the use of social media on patients changing dental practice: a web-based questionnaire study. BMC Oral Health 2023 Jun 6;23(1):365. [doi: 10.1186/s12903-023-03078-9] [Medline: 37277765]
10. MacDonald J, Sohn S, Ellis P. Privacy, professionalism and Facebook: a dilemma for young doctors. Med Educ 2010 Aug;44(8):805-813. [doi: 10.1111/j.1365-2923.2010.03720.x] [Medline: 20633220]
11. Malik A, Heyman-Schrum C, Johri A. Use of Twitter across educational settings: a review of the literature. Int J Educ Technol High Educ 2019 Dec;16(1):1-22. [doi: 10.1186/s41239-019-0166-x]
12. Forgie SE, Duff JP, Ross S. Twelve tips for using Twitter as a learning tool in medical education. Med Teach 2013;35(1):8-14. [doi: 10.3109/0142159X.2012.746448] [Medline: 23259608]
13. Madathil KC, Rivera-Rodriguez AJ, Greenstein JS, Gramopadhye AK. Healthcare information on YouTube: a systematic review. Health Informatics J 2015 Sep;21(3):173-194. [doi: 10.1177/1460458213512220] [Medline: 24670899]
14. Simsek H, Buyuk SK, Cetinkaya E. YouTube as a source of information on oral habits. J Indian Soc Pedod Prev Dent 2020;38(2):115-118. [doi: 10.4103/JISPPD.JISPPD_357_19] [Medline: 32611855]
15. Guo J, Yan X, Li S, Van der Walt J, Guan G, Mei L. Quantitative and qualitative analyses of orthodontic-related videos on YouTube. Angle Orthod 2020 May 1;90(3):411-418. [doi: 10.2319/082019-542.1] [Medline: 33378439]
16. Ooi HL, Kelleher MGD. Instagram dentistry. Prim Dent J 2021 Mar;10(1):13-19. [doi: 10.1177/2050168420980964] [Medline: 33722130]
17. Patil P, Pawar AM, Agustin Wahjuningrum D, Maniangat Luke A, Reda R, Testarelli L. Connectivity and integration of Instagram use in the lives of dental students and professionals: a country-wide cross-sectional study using the InstaAA Questionnaire. J Multidiscip Healthc 2023;16:2963-2976. [doi: 10.2147/JMDH.S430973] [Medline: 37818291]
18. Buyuk SK, Alpaydın MT, Imamoglu T. Social media and orthodontics: an analysis of orthodontic-related posts on Instagram. Aust Orthod J 2020 Jan 1;36(2):153-159. [doi: 10.21307/aoj-2020-017]
19. Fraticelli L, Smentek C, Tardivo D, et al. Characterizing the content related to oral health education on TikTok. Int J Environ Res Public Health 2021 Dec 16;18(24):13260. [doi: 10.3390/ijerph182413260] [Medline: 34948869]
20. Kılınç DD. Is the information about orthodontics on Youtube and TikTok reliable for the oral health of the public? A cross sectional comparative study. J Stomatol Oral Maxillofac Surg 2022 Oct;123(5):e349-e354. [doi: 10.1016/j.jormas.2022.04.009] [Medline: 35470122]
21. Kurian N, Varghese IA, Cherian JM, et al. Influence of social media platforms in dental education and clinical practice: a cross-sectional survey among dental trainees and professionals. J Dent Educ 2022 Aug;86(8):1036-1042. [doi: 10.1002/jdd.12914] [Medline: 35257384]
22. Hamasha AAH, Alghofaili N, Obaid A, et al. Social media utilization among dental practitioner in Riyadh, Saudi Arabia. Open Dent J 2019 Feb 28;13(1):101-106. [doi: 10.2174/1874210601913010101]
23. Aboalshamat K, Alkiyadi S, Alsaleh S, et al. Attitudes toward social media among practicing dentists and dental students in clinical years in Saudi Arabia. Open Dent J 2019 Mar 28;13(1):143-149. [doi: 10.2174/1874210601913010143]
24. Maharani DA, El Tantawi M, Yoseph MG, Rahardjo A. The use of internet platforms for oral health information and associated factors among adolescents from Jakarta: a cross sectional study. BMC Oral Health 2021 Jan 7;21(1):22. [doi: 10.1186/s12903-020-01387-x] [Medline: 33413315]
25. Althunayan A, Alsalhi R, Elmoazen R. Role of social media in dental health promotion and behavior change in Qassim province, Saudi Arabia. Int J Med Health Res 2018;4(2):98-103 [FREE Full text]
26. 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]
27. Farrokhi F, Ghorbani Z, Farrokhi F, Namdari M, Salavatian S. Social media as a tool for oral health promotion: a systematic review. PLoS ONE 2023;18(12):e0296102. [doi: 10.1371/journal.pone.0296102] [Medline: 38113216]
28. Sivrikaya EC, Yilmaz O, Sivrikaya P. Dentist-patient communication on dental anxiety using the social media: a randomized controlled trial. Scand J Psychol 2021 Dec;62(6):780-786. [doi: 10.1111/sjop.12769] [Medline: 34333783]
29. Farsi D. Social media and health care, part I: literature review of social media use by health care providers. J Med Internet Res 2021 Apr 5;23(4):e23205. [doi: 10.2196/23205] [Medline: 33664014]
30. Ghandhi D, Bodani N, Lal A, et al. Evaluation of social media usage by dental practitioners of Pakistan for professional purposes - a cross-sectional study. Clin Cosmet Investig Dent 2022;14:245-252. [doi: 10.2147/CCIDE.S374111] [Medline: 35975279]
31. Henry RK, Molnar A, Henry JC. A survey of US dental practices' use of social media. J Contemp Dent Pract 2012 Mar 1;13(2):137-141. [doi: 10.5005/jp-journals-10024-1109] [Medline: 22665737]

32. Bahabri RH, Zaidan AB. The impact of social media on dental practice promotion and professionalism amongst general dental practitioners and specialists in KSA. J Taibah Univ Med Sci 2021 Jun;16(3):456-460. [doi: 10.1016/j.jtumed.2020.12.017] [Medline: 34140874]
33. Number of internet and social media users worldwide as of July 2024 (in billions). Statista. URL: https://www.statista.com/statistics/617136/digital-population-worldwide/ [accessed 2024-12-17]
34. Most popular social networks worldwide as of April 2024, by number of monthly active users (in millions). Statista. URL: https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/ [accessed 2024-12-17]
35. Social media in the Philippines - statistics & facts. Statista. URL: https://www.statista.com/topics/6759/social-media-usage-in-the-philippines/#topicOverview [accessed 2024-12-17]
36. Al-Khalifa KS, Al-Swuailem AS, AlSheikh R, et al. The use of social media for professional purposes among dentists in Saudi Arabia. BMC Oral Health 2021 Jan 12;21(1):26. [doi: 10.1186/s12903-021-01390-w] [Medline: 33435911]
37. Jamkhande A, Pangarikar AB, Khairnar MR, Gavali N, Patil Y, Wadgave U. Social media use in dental practice by dental practitioners of India: a cross-sectional survey. J Clin Diagn Res 2022;16(6). [doi: 10.7860/JCDR/2022/56404.16533]
38. Makkar N, Lawrence D, Veeresh D, Banik A, Kumar A, Mehar AS. Assessment of social media usage and its influence among dental faculty members in Davangere city – a cross sectional survey. J Indian Assoc Public Health Dent 2018;16(3):246. [doi: 10.4103/jiaphd.jiaphd_143_17]
39. Kamarudin Y, Mohd Nor NA, Libamin AC, et al. Social media use, professional behaviors online, and perceptions toward e-professionalism among dental students. J Dent Educ 2022 Aug;86(8):958-967. [doi: 10.1002/jdd.12912] [Medline: 35247218]
40. Almaiman S, Bahkali S, Al Farhan A, Bamuhair S, Househ M, Alsurimi K. The prevalence of using social media among healthcare professionals in Saudi Arabia: a pilot study. Stud Health Technol Inform 2015;213:263-266. [Medline: 26153010]
41. 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. [doi: 10.2196/mededu.6232] [Medline: 27731855]
42. Snyman L, Visser JH. The adoption of social media and social media marketing by dentists in South Africa. SADJ 2014 Jul;69(6):258. [Medline: 26548200]
43. Kemp S. Digital 2024: the Philippines. Datareportal. 2024 Feb 21. URL: https://datareportal.com/reports/digital-2024-philippines [accessed 2024-12-17]
44. Zareban I, Rahmani A, Allahqoli L, et al. Effect of education based on trans-theoretical model in social media on students with gingivitis; a randomized controlled trial. Health Educ Health Prom 2022;10(1):1-7 [FREE Full text]
45. Scheerman JFM, Hamilton K, Sharif MO, Lindmark U, Pakpour AH. A theory-based intervention delivered by an online social media platform to promote oral health among Iranian adolescents: a cluster randomized controlled trial. Psychol Health 2020 Apr;35(4):449-466. [doi: 10.1080/08870446.2019.1673895] [Medline: 31621423]
46. Scribante A, Gallo S, Bertino K, Meles S, Gandini P, Sfondrini MF. The effect of chairside verbal instructions matched with Instagram social media on oral hygiene of young orthodontic patients: a randomized clinical trial. Appl Sci (Basel) 2021;11(2):706. [doi: 10.3390/app11020706]
47. Zotti F, Dalessandri D, Salgarello S, et al. Usefulness of an app in improving oral hygiene compliance in adolescent orthodontic patients. Angle Orthod 2016 Jan;86(1):101-107. [doi: 10.2319/010915-19.1] [Medline: 25799001]
48. Al-Silwadi FM, Gill DS, Petrie A, Cunningham SJ. Effect of social media in improving knowledge among patients having fixed appliance orthodontic treatment: a single-center randomized controlled trial. Am J Orthod Dentofacial Orthop 2015 Aug;148(2):231-237. [doi: 10.1016/j.ajodo.2015.03.029] [Medline: 26232831]
49. Al-Nahdi S, Alanzi T, Alhodaib H. Use of social media by healthcare providers in saudi arabia: reasons and usefulness for their professional development. Presented at: 19th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES); Oct 16-19, 2020; Xuzhou, China. [doi: 10.1109/DCABES50732.2020.00091]
50. Terrasse M, Gorin M, Sisti D. Social media, e-health, and medical ethics. Hast Cent Rep 2019 Jan;49(1):24-33. [doi: 10.1002/hast.975] [Medline: 30790306]

Abbreviations

STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by A Mavragani; submitted 05.09.24; peer-reviewed by J Hunsrisakhun, K Sadasiva; revised version received 19.11.24; accepted 20.11.24; published 03.01.25.

Please cite as:
Acosta JM, Detsomboonrat P, Pisarnturakit PP, Urwannachotima N
The Use of Social Media on Enhancing Dental Care and Practice Among Dental Professionals: Cross-Sectional Survey Study
JMIR Form Res 2025;9:e66121
URL: https://formative.jmir.org/2025/1/e66121
doi:10.2196/66121

Artificial Intelligence–Powered Training Database for Clinical Thinking: App Development Study

Heng Wang[1*], MEd; Danni Zheng[1*], PhD; Mengying Wang[2], MEng; Hong Ji[2], MEng; Jiangli Han[3], MD; Yan Wang[4], MD; Ning Shen[5], MD; Jie Qiao[6], MD

[1]Education Department, Peking University Third Hospital, Beijing, China

[2]Information Management and Big Data Center, Peking University Third Hospital, Beijing, China

[3]Department of Cardiology, Peking University Third Hospital, Beijing, China

[4]Department of Obstetrics and Gynecology, Peking University Third Hospital, Beijing, China

[5]Department of Pulmonary and Critical Care Medicine, Peking University Third Hospital, Beijing, China

[6]Centre for Reproductive Medicine, Department of Obstetrics and Gynecology, Peking University Third Hospital, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Jie Qiao, MD
Centre for Reproductive Medicine, Department of Obstetrics and Gynecology, Peking University Third Hospital, Beijing, China

Abstract

Background: With the development of artificial intelligence (AI), medicine has entered the era of *intelligent medicine*, and various aspects, such as medical education and talent cultivation, are also being redefined. The cultivation of clinical thinking abilities poses a formidable challenge even for seasoned clinical educators, as offline training modalities often fall short in bridging the divide between current practice and the desired ideal. Consequently, there arises an imperative need for the expeditious development of a web-based database, tailored to empower physicians in their quest to learn and hone their clinical reasoning skills.

Objective: This study aimed to introduce an app named "XueYiKu," which includes consultations, physical examinations, auxiliary examinations, and diagnosis, incorporating AI and actual complete hospital medical records to build an online-learning platform using human-computer interaction.

Methods: The "XueYiKu" app was designed as a contactless, self-service, trial-and-error system application based on actual complete hospital medical records and natural language processing technology to comprehensively assess the "clinical competence" of residents at different stages. Case extraction was performed at a hospital's case data center, and the best-matching cases were differentiated through natural language processing, word segmentation, synonym conversion, and sorting. More than 400 teaching cases covering 65 kinds of diseases were released for students to learn, and the subjects covered internal medicine, surgery, gynecology and obstetrics, and pediatrics. The difficulty of learning cases was divided into four levels in ascending order. Moreover, the learning and teaching effects were evaluated using 6 dimensions covering systematicness, agility, logic, knowledge expansion, multidimensional evaluation indicators, and preciseness.

Results: From the app's first launch on the Android platform in May 2019 to the last version updated in May 2023, the total number of teacher and student users was 6209 and 1180, respectively. The top 3 subjects most frequently learned were respirology (n=606, 24.1%), general surgery (n=506, 20.1%), and urinary surgery (n=390, 15.5%). For diseases, pneumonia was the most frequently learned, followed by cholecystolithiasis (n=216, 14.1%), benign prostate hyperplasia (n=196, 12.8%), and bladder tumor (n=193, 12.6%). Among 479 students, roughly a third (n=168, 35.1%) scored in the 60 to 80 range, and half of them scored over 80 points (n=238, 49.7%). The app enabled medical students' learning to become more active and self-motivated, with a variety of formats, and provided real-time feedback through assessments on the platform. The learning effect was satisfactory overall and provided important precedence for establishing scientific models and methods for assessing clinical thinking skills in the future.

Conclusions: The integration of AI and medical education will undoubtedly assist in the restructuring of education processes; promote the evolution of the education ecosystem; and provide new convenient ways for independent learning, interactive communication, and educational resource sharing.

(JMIR Form Res 2025;9:e58426) doi:10.2196/58426

KEYWORDS

artificial intelligence; clinical thinking ability; virtual medical records; distance education; medical education; online learning

Introduction

Clinical thinking refers to the ability of doctors to understand, diagnose, and treat diseases through analysis, synthesis, judgment, reasoning, and other thinking activities by using their medical knowledge, skills, and experience [1]. Clinical reasoning skills include a series of strategies such as correlating information, conducting a comprehensive analysis to form one or more diagnostic hypotheses, weighing the risks and benefits of diagnostic tests and treatments, as well as formulating a reasonable diagnosis and treatment plan [2]. This detailed process includes collecting and evaluating clinical information, selecting diagnostic tests, evaluating test results, developing diagnostic hypotheses, and weighing treatment options and risks [3].

Teaching clinical thinking ability is a challenging task even for experienced clinical teachers. As clinical knowledge is the basis of clinical thinking, the cultivation of clinical thinking needs to go through the links of guidance, deliberate practice, and feedback, and it takes a long time to cultivate efficient clinical thinking ability. The level of clinical thinking of residents is directly related to the safety and quality of medical treatment [4]. Therefore, we should constantly improve the understanding of the clinical thinking training and improve the training quality of residents.

Therefore, the cultivation of clinical thinking ability is a necessary condition for the transition from medical students to clinicians and is currently the key point and difficulty among resident training programs. Among the training processes, the most effective method is based on real or virtual case training, combined with feedback and reflection to continuously improve. Previous studies have shown that clinical thinking ability is scarce among physicians, while offline training cannot meet the gap between the present reality and the ideal requirements [5]. Therefore, there is an urgent need for the development of a web-based database to help physicians learn and exercise clinical thinking ability.

Artificial intelligence (AI) is a novel technology that can be utilized to research and develop the theories, methods, technologies, and applications used to simulate, extend, and expand human intelligence. AI belongs to the interdisciplinary field of natural science, social science, and technological science [6]. AI has a history of more than 70 years, and its application in the medical field mainly focuses on five aspects: medical imaging, auxiliary diagnosis, drug research and development, health management, and disease prediction.

There have been several reports and studies on the use of AI in medical teaching and training of clinical thinking ability. For example, Diagnostic Reasoning (DxR) Clinician (DxR Development Group, Inc.) was developed in 1994 by Myers and Dorsey in the United States to train clinical thinking among students [7]. In 2006, the BP network neural algorithm in AI was applied to teach quality monitoring instead of manually teaching daily monitoring [8]. In 2011, Knewton, an adaptive education platform, offered personalized learning services by extracting learning data from students. In 2017, Tsinghua University launched a study system combined with AI and full quantitative reality virtual technology. In this system, patients' image data can be converted into a holographic three-dimensional anatomical structure of the human body and mapped in the virtual space [9]. In 2020, a competency-based learning and assessment system (CBLAS) for residency training was designed to provide resident physicians with clinical assessments and learning in order to enhance the learning of trainees and reduce the burden of assessments [10]. However, in the era of AI, there are still issues that exist in medical teaching, such as the lack of general education, the lack of classroom interaction, and the lag in curriculum construction [11].

On this basis, the present study introduces an app named "XueYiKu," which includes consultations, physical examinations, auxiliary examinations, and diagnosis, by incorporating AI and actual complete hospital medical records to build an online-learning platform using human-computer interaction. This application could cultivate students' ability to solve clinical problems as well as improve teaching quality and efficiency, thereby achieving the sharing of high-quality teaching resources.

Methods

Design and Study Population

The "XueYiKu" app was designed as a contactless, self-service, trial-and-error system application based on actual complete hospital medical records and natural language processing technology to build an online-learning platform using human-computer interaction. Training participants included medical undergraduates, medical graduate students, residents, chief residents, and attending doctors. The purpose of the app was to improve clinical thinking ability through independent study accompanied by real-time feedback. The XueYiKu app was launched on the Android platform on May 14, 2019, and on the IOS platform on August 3, 2019. After 27 updates, the last version of the XueYiKu app was updated on January 12, 2022.

Data Collection

The general database of the XueYiKu app was based on a collection of real medical records and data from general teaching, including a standard clinical question database, virtual medical record database, and student learning record database. Using information extraction technology in natural language processing from real medical records, the XueYiKu app could poststructure the content from real electronic medical records, generate questions, and match answers online through machine learning. In the teaching cases, the automatically extracted answers matched 90% of the original medical records. The structured cases in the electronic medical record system of the hospital were then screened by a teaching team from Peking University Third Hospital, and appropriate medical records were selected and reviewed by the teaching team. Once the teaching case was determined, relevant inspection results were also extracted, including inspections, auxiliary inspections, and ultrasound, imaging, and pathology findings.

Outcome Measures

Clinical thinking ability involves extremely high requirements for the objectivity and standardization of evaluation. Hence, a multidimensional automatic evaluation system is required. With reference to the residents' standardized training report, the whole process of student learning was recorded systematically, and clinical thinking was evaluated in multiple dimensions according to the learning situation of the student, including preciseness, logic, systematicness, agility, knowledge expansion, and multidimensional and comprehensive application (Table 1). Based on the scores of students in various dimensions of learning, a radar chart of clinical thinking ability is drawn to assess the compliance of medical students' clinical thinking ability.

Table . Seven elements for clinical thinking evaluation in the XueYiKu app.

Elements	Contents
Preciseness	Comprehensive evaluation of diagnosis and differential diagnosis
Logic	The logical sequence evaluation of inquiry, physical examination, auxiliary examination, and diagnosis
Systematicness	Comprehensive evaluation of the completion rate of inquiry, physical examination, auxiliary examination, diagnosis, and differential diagnosis
Agility	Comprehensive evaluation of the time of inquiry, physical examination, auxiliary examination, diagnosis, and differential diagnosis
Knowledge expansion	Accuracy and logical evaluation of diagnosis and differential diagnosis
Multidimensional evaluation indicators	Comprehensive evaluation of diagnosis and treatment programs
Comprehensive application	Comprehensive application ability of basic and clinical knowledge (for evaluation of level 4 individually)

Evaluation and Feedback

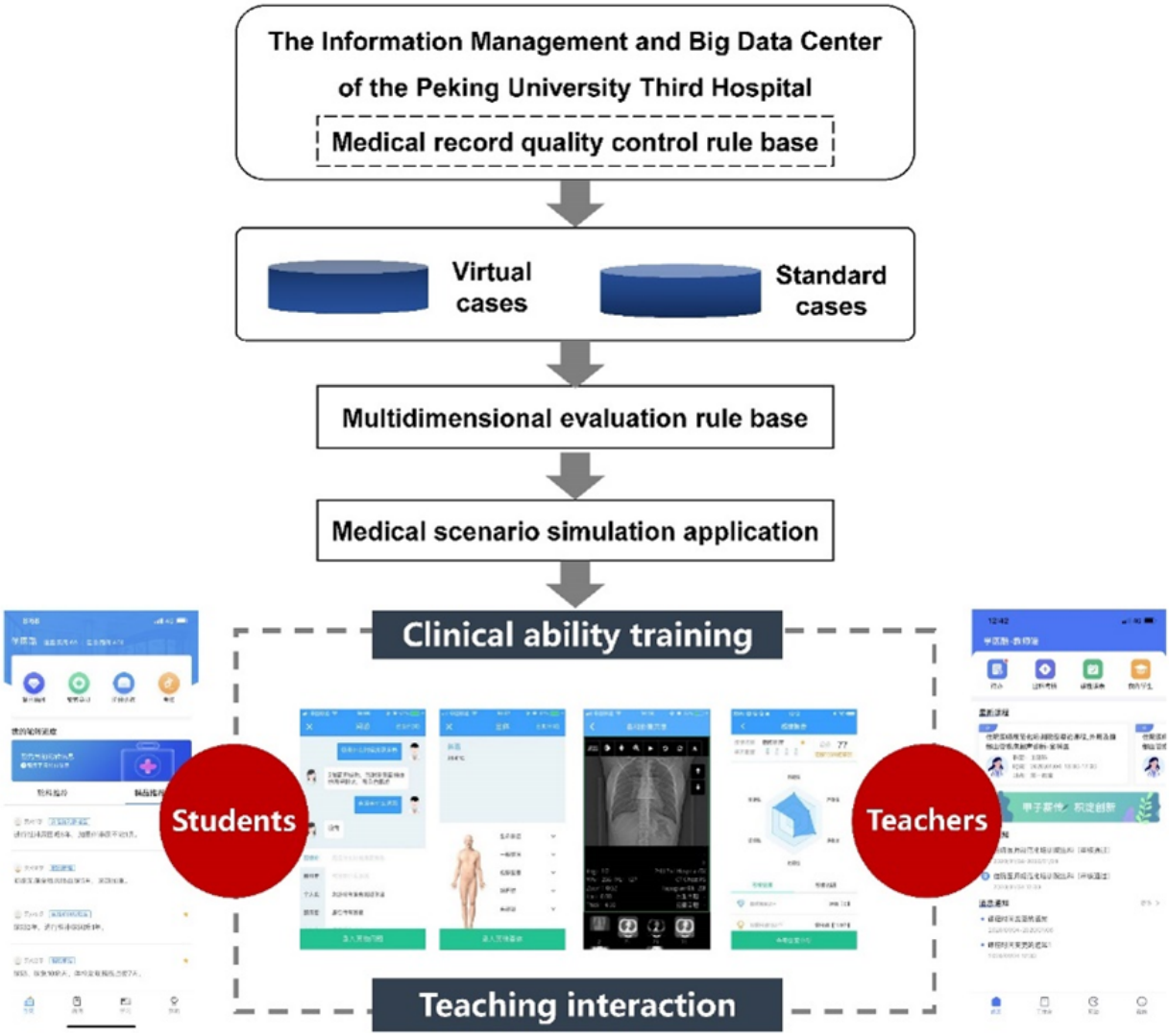
We recruited all students in Peking University Hospital to evaluate the teaching effect of the XueYiKu app, including medical undergraduates, medical graduate students, residents, chief residents, and attending doctors. Both the questionnaire method and interview method were utilized. Web-based surveys, developed by the study team, were conducted upon completion of the case study so that students could fill in and evaluate their learning effect. Through interviews with relevant teachers and experts, the common clinical path extracted using AI technology was determined. Meanwhile, the range of clinical thinking ability evaluation was confirmed according to the requirements of different training stages. The questionnaire method was used to investigate the following two factors: (1) the current status of clinical thinking training regarding the ability of medical students and residents to clarify the problems in clinical thinking training and improve the content of the database; (2) the experience of medical students and residents regarding the use of the case database and self-evaluation of learning effects to complete the effect evaluation of the XueYiKu app.

Data Processing

Details of the virtual case system and AI teaching development based on actual complete hospital medical records and natural language processing technology have been described previously [12]. A flow chart of the program design from data collection to interaction between students and teachers is shown in Figure 1. Case extraction was performed at a hospital's case data center, and the best-matching cases were extracted through natural language processing, word segmentation, synonym conversion, and sorting. A standard clinical questioning data module, virtual case data module, and student learning difficulty module were established to achieve simulation. Students can view the objective examination and inspection data of actual cases, including details of the consultation and physical examination, and automatically provide their learning response via a multidimensional evaluation system. In order to assess the changes in students' clinical thinking after using the XueYiKu app, 15 medical graduate students were subjected to two simulation tests before and after learning through the virtual case system. The tests, which included the full-process case examination of cases having the same difficulty level, examined core clinical thinking test points such as consultation, physical examination, and disposal, and generated multidimensional evaluation indicators (rigor, logic, system, agility, and knowledge expansion). Thus, a complete and credible evaluation system was developed.

For statistical analysis, standard descriptive statistics were used to describe the study population: continuous variables were presented as the mean and SD, and categorical variables were presented as percentages.

Figure 1. Flow chart of the program design from data collection to interaction between students and teachers in the XueYiKu app.



Ethical Considerations

This study was reviewed and approved by the Peking University Third Hospital Medical Science Research Ethics Committee (No. IRB00006761-M2022063). Invitations were dispatched to prospective participants via the XueYiKu app, accompanied by a comprehensive informed consent form outlining the nature of the investigation. Participants were clearly notified that the interviews would be audio-recorded solely for analytical purposes and would remain exclusively accessible to the research team. Furthermore, they were assured of their right to withdraw their consent at any juncture during the interview process. All information gathered about the participants was kept private and confidential. All study data were deidentified.

Results

Overview

After concealment processing, more than 400 teaching cases covering 65 kinds of diseases were released for students to learn from, and the subjects covered internal medicine, surgery, gynecology and obstetrics, and pediatrics (Table 2). The difficulty in learning cases was divided into 4 levels in ascending order: level 1 provided all score items for students to choose from; except for items from level 1, level 2 provided additional interference items; compared with level 2, level 3 displayed some score items and all interference items to enhance difficulty; for level 4, there was neither a score item nor an interference item, and the system would match the results from the students' voice or manual input of the questions.

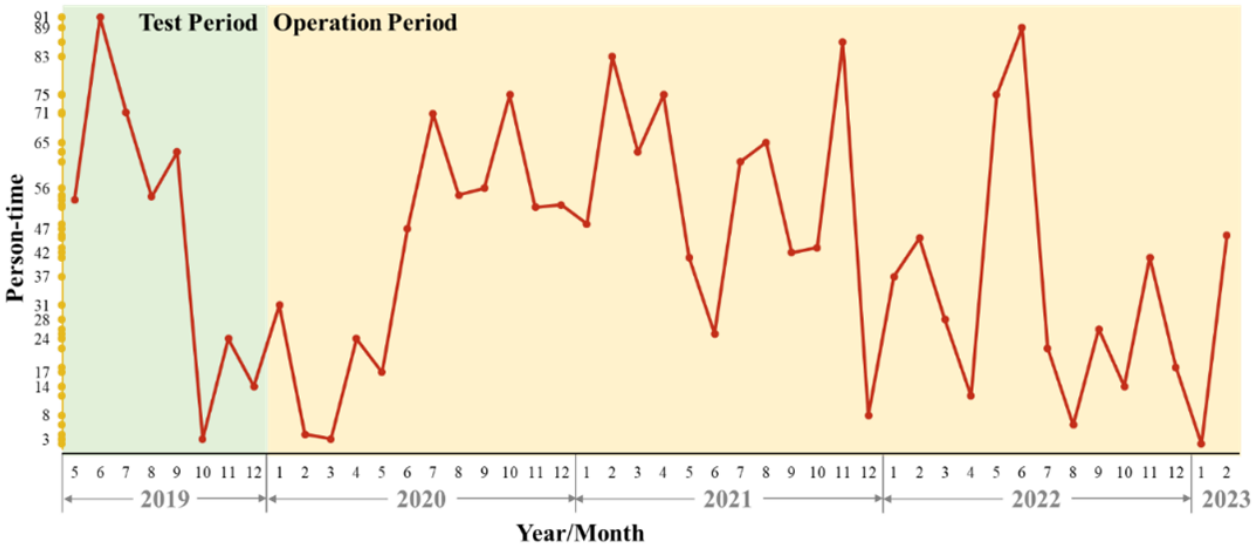
Table . Categories of subjects and diseases for learning and evaluation in the database of the XueYiKu app.

Subjects	Category	Diseases
Internal Medicine	20	Coronary heart disease, heart failure, arrhythmia, hypertension, pneumonia, chronic airway inflammatory disease, pleural effusion, pulmonary embolism, upper gastrointestinal bleeding, pancreatitis, inflammatory bowel disease, ascites, nephrotic syndrome, acute kidney injury, diabetes, hyperthyroidism, leukemia, anemia, systemic lupus erythematosus, rheumatoid arthritis
Surgery	20	Inguinal hernia, appendicitis, gallbladder stones, bile duct stones, stomach cancer, colon cancer, rectal cancer, liver cancer, pancreatic cancer, lung cancer, pneumothorax, pituitary tumor, coronary heart disease, benign prostatic hyperplasia, bladder tumor, acute respiratory distress syndrome, cervical myelopathy disease, lumbar disc herniation, knee osteoarthritis, ankle fracture
Gynecology and obstetrics	15	Ectopic pregnancy, uterine fibroids, ovarian tumors, dysfunctional uterine bleeding, endometrial cancer, cervical cancer, ovarian cancer, hypertension during pregnancy (including gestational hypertension, preeclampsia, eclampsia, chronic hypertension, and chronic hypertension combined with preeclampsia), gestational diabetes or diabetes (impaired glucose tolerance) combined with pregnancy, prenatal hemorrhage (including placenta previa and abruption of the placenta)
Pediatrics	10	Neonatal hyperbilirubinemia, neonatal respiratory distress syndrome, neonatal sepsis, neonatal hypoxic-ischemic encephalopathy, pneumonia, diarrhea, nutritional diseases, nervous system disease, Kawasaki disease, and kidney disease

App Use Status

From the first launch on the Android platform on May 14, 2019, to the last version of the XueYiKu app updated in May 2023, the total number of teacher and student users was 6209 and 1180, respectively. Figure 2 represents the use status of the XueYiKu app during the test period and operation period. In total, there were 3224 person-times in the last 3 years, with an average of 771 person-times per year. From 2019 to 2021, the person-times logged in the application were the highest in 2019, at 1032 (125.4 on average per month), and increased from 646 in 2020 (53.8 on average per month) to 994 in 2021 (76.6 on average per month).

Figure 2. Changes in the number of student users of the XueYiKu app from May 2019 to February 2023.



We calculated the frequency at which subjects and related diseases were learned by students, and the top 10 most frequently learned subjects and diseases are listed in Table 3 in descending order. For the second discipline, the most frequently learned subjects were respirology (n=606, 62.2% in internal medicine), general surgery (n=506, 48.8% in surgery), and urinary surgery (n=390, 37.6% in surgery). For diseases, pneumonia was the most frequently learned, with 452 person-times (58.4% in internal medicine), followed by cholecystolithiasis (n=216, 28.4% in surgery), benign prostate hyperplasia (n=196, 25.8% in surgery), and bladder tumor (n=193, 25.4% in surgery).

Table . The top 10 most frequently learned subjects and diseases in the XueYiKu app.

	Person-time	Percentage (%)
Subjects[a]		
Internal Medicine	975	38.8
Respirology	606	62.2
Cardiology	201	20.6
Gastroenterology	88	9
Endocrinology	80	8.2
Surgery	1036	41.2
General surgery	506	48.8
Urinary surgery	390	37.6
Thoracic surgery	140	13.5
Gynecology and obstetrics	338	13.5
Gynecology	175	51.8
Obstetrics	163	48.2
Pediatrics	163	6.5
Diseases[a]		
Internal Medicine	774	50.4
Pneumonia	452	58.4
Coronary heart disease	99	12.8
Lung cancer	79	10.2
Arrhythmia	76	9.8
Chronic airway inflammation	68	8.8
Surgery	761	49.6
Cholecystolithiasis	216	28.4
Benign prostate hyperplasia	196	25.8
Bladder tumor	193	25.4
Appendicitis	79	10.4
Inguinal hernia	77	10.1

[a]The top 10 subjects and diseases were listed in descending order due to quantitative limitations.

Evaluation of the App and its Learning Effects

To evaluate the learning effect of the XueYiKu app, the scores of 479 students were analyzed. On the basis of a full score of 100, 15% (n=73) of students scored less than a passing grade of 60 points, roughly a third (n=168, 35.1%) had a score in the 60 to 80 range, and half of them had a score of over 80 points (n=238, 49.7%). A radar chart (Figure 3) with specific scores (Table 4) was presented considering multiple dimensions to evaluate the learning situation of students: students who learned cases with difficulty levels 3 - 4 achieved excellent learning outcomes regarding preciseness (83.4 for level 3), multiple dimensions (87.4 for level 4), and knowledge expansion (86.5 for level 3). However, students who learned cases with a difficulty of level 4 did not score as high as the other students in systematicness (34.1 for level 4), agility (30.1 for level 4), and logic (21.1 for level 4). In contrast, students who learned from cases with level 3 difficulty had positive effects in many areas (86.5 in knowledge expansion and 83.4 in preciseness). In fact, there was a high level of satisfaction with knowledge expansion and preciseness in all learners, but it is worth noting

that all levels were completely negative about systematicness, agility, and logic.

Figure 3. Radar chart with specific scores of clinical thinking evaluation for the XueYiKu app by difficulty in learning cases.

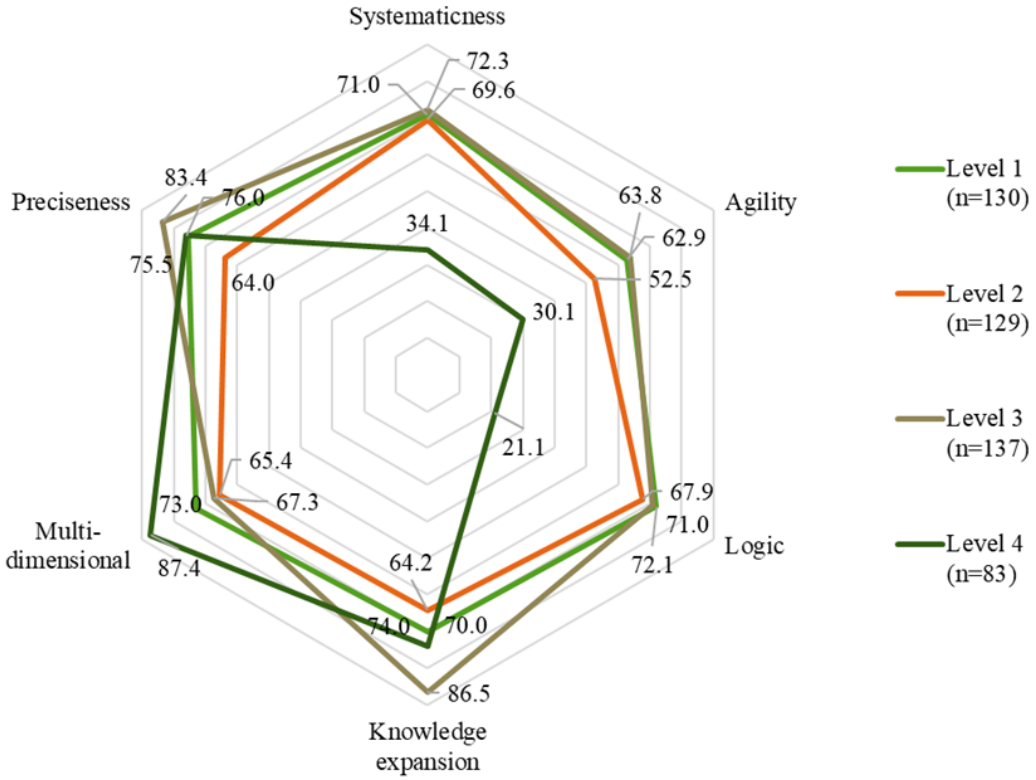


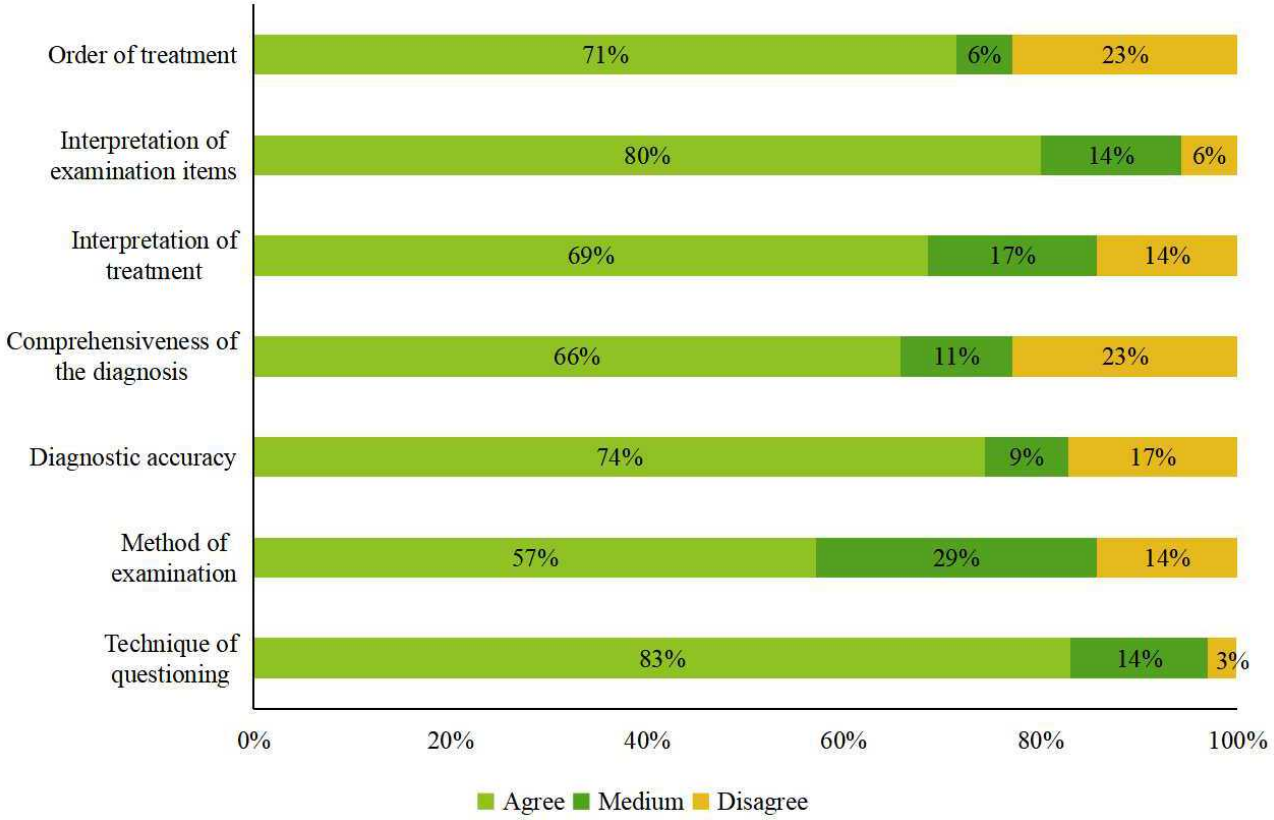
Table . Scores of clinical thinking for the XueYiKu app by the difficulty in learning cases and 6 dimensions according to the learning situation of the student.

	Level 1 (n=130)	Level 2 (n=129)	Level 3 (n=137)	Level 4 (n=83)
Systematicness	71.0	69.6	72.3	34.1
Agility	62.9	52.5	63.8	30.1
Logic	72.1	67.9	71.0	21.1
Knowledge expansion	70.0	64.2	86.5	74.0
Multidimensional evaluation indicators	73.0	65.4	67.3	87.4
Preciseness	75.5	64.0	83.4	76.0

Evaluation of the App and Its Learning Effects

This section provides information regarding the 70 students' satisfaction of the XueYiKu app and their self-assessment of its learning effect (Figure 4). Overall, most students were satisfied with the app. The degree of satisfaction kept increasing considering the app function, the personal learning effect, and the case quality. For the self-assessment of learning outcomes, all sections received a rating of more than satisfactory, but the method of examination was rated the least satisfactory. In other sections, the technique of questioning received the highest satisfaction rating, followed by the interpretation of examination items, with both categories receiving more than 80% positive feedback. In contrast, the diagnostic accuracy and the order and interpretation of treatment ranked next in order (~70%), and the other factors were at approximately 60%. Most of the dissatisfaction was rated for the order of treatment and the comprehensiveness of the diagnosis, which were responsible for more than 20% of the negative ratings. Except for the questioning technique, the proportion of negative ratings was more pronounced in all sections. Additionally, for the method of examination, about 30% rated "medium," which differed greatly from other sections' ratings.

Figure 4. Self-assessment and satisfaction of 70 students towards the learning outcomes for the XueYiKu app.



Feedback on the App and Its Teaching Effects

According to the feedback, there were some issues associated with the app that are worthy of attention. First, the original login method should be simplified, as well as that for the online case communication and sharing section. Additionally, discussion and comment functions should be added. Regarding data processing, the AI answers could optimize the fuzzy search, and the system could be better connected to the residency system so that students can check and manage the patient's detailed medical record information. Regarding the case corpus, there is still no section in the app to read images. In addition, not only do the users of the app want to obtain clearer guidance or feedback in future updates, but it would also be beneficial if special departments fulfilled the auxiliary examination and clinical guidelines. Finally, some of the cases included in the app are so simple that they are only suited for undergraduate medical students; it would be beneficial to add cases that are rare, complex, or even life-threatening, and to include the postoperative complications, judgment, treatment, and other information relevant to the case.

Discussion

Principal Findings

Clinical thinking training for medical students is very important; however, few design cases rely on IT for self-learning. The current study uses AI and big data platforms to establish a virtual learning platform based on real clinical cases. The XueYiKu app can assist in the development of medical students' clinical thinking skills by allowing them to perform consultation, physical examination, auxiliary examination, diagnosis, and treatment on virtual patients; the app can also design real cases with different difficulties to meet the needs of residents' clinical thinking training at different stages. Through the evaluation of students' experiences with the app, this study concludes that medical students' learning can become more active and self-motivated, using a variety of formats, with real-time feedback through the platform's assessment. The learning effect is also satisfactory overall and provides important methods and ideas for establishing scientific models and methods for assessing clinical thinking skills in the future.

Comparison to Prior Work

Analysis of the data shows that the number of case studies was positively correlated with the number of residents in the specialty. The top 4 diseases most frequently learned included pneumonia, cholecystolithiasis, benign prostate hyperplasia, and bladder tumor, all of which are common, require adequate diagnosis and differential diagnosis, and meet the basic goal of clinical thinking training at the residency stage. The platform contains 65 common diseases and 400 study cases covering internal medicine, surgery, obstetrics and gynecology, and pediatrics. The design is reasonable and can meet the basic learning needs of residents in various specialties. In the 1990s, DxR Clinician was developed by DxR Development Group, Inc. [7] in the United States, and it was later introduced and translated into Chinese by the China Union Medical University Press in 2009 [13]. However, the use of this software in China is still in its infancy, with few relevant literature reports. A previous study showed that the cases included in the DxR software were virtual cases and were limited by the purchase [7]. Students needed to complete the study on the computer, and the average time for students to become familiar with the

software was as long as 2 hours. In contrast, the types of diseases and the number of cases covered in the XueYiKu app, as well as the ease of use, are advantages. In addition, the framework of this platform is reasonable, as new case types of diseases can be added according to students' needs, and the user interface is simple and fast. Therefore, this app is more suitable for residents' clinical thinking training in the global internet and postepidemic era.

There is a decline in the training and formation of clinical thinking in residents. Based on this, the real case base used in this app was designed to present 4 levels of difficulty for the same case. Evaluation and feedback in the process of clinical thinking training formation and refinement are also important. As a result, students were evaluated in multiple dimensions according to the learning situation. Analysis of the results showed that the learning process meets the clinical diagnosis and treatment reality. For first- to second-year residents, level 2 - 3 difficulty was the most appropriate, as illustrated by their better performance in knowledge expansion and logic, among other factors. The clinical learning process is in line with clinical practice and meets the learning needs of residents. Specific to the learning process at different levels of difficulty, the primary difficulty allowed for a complete diagnosis based on the completed clinical information, and as the difficulty level increased, there was an advancement of the learning model; the results of the 4 levels of difficulty were in line with the development of the resident's clinical thinking, realizing the intended purpose of the app. One similar app currently in use is Brianly, a medical tutoring and social networking platform that provides personalized medical learning tutoring through AI algorithms [14]. In Tsinghua University, "artificial intelligence + virtual reality" within Brianly is used for clinical training, which allows direct viewing of the anatomical details of a patient's real body structure in a virtual space, as well as virtual surgery. However, there is currently no learning software for medical students that is designed to match the growth and learning process, and that provides a difficulty gradient based on the clinical thinking ability. In this context, our current app will continue to develop, improve, and provide feedback related to the 4 levels of difficulty to be better used for the clinical thinking training of senior residents.

Based on current usage, the highest number of uses per capita was usually found during the admission and start of residency, that is, from September to February each year. The survey results show that students had a high level of satisfaction with the consultation aspect of the app. The case study on the app allows for both a generic history-taking process and a structured questioning for a particular disease, which is inextricably linked to the teaching case base from clinical consultations of real cases. However, there are some differences between the examination and the actual situation, and the contents of the specialist examination and the physical examination for a particular disease should be further optimized to be more concise and focused.

Strengths and Limitations

One of the greatest advantages of our app is the inclusion of real cases, which could provide a new model of clinical thinking training by building a system to comprehensively assess the "clinical competence" of residents at different stages. For students, the strongest advantage is that they can customize clinical competency training according to their needs without the limitations of time and space. For app designers, the XueYiKu app is inherently user-friendly and sustainable, which allows the addition of cases in bulk based on the existing framework.

In addition to the benefits of the app, there are several limitations that could be further improved. First, current cases and diseases could be further expanded to cover more subjects, which is the next step of development. Second, in terms of use, students currently use the app independently, while the frequency of use is mainly based on students' interests combined with required course study and clinical rotation. Third, current evaluation results are almost subjective evaluations by humans. Thus, the evaluation results may be biased.

Conclusion

The emergence of virtual network learning systems such as the XueYiKu app can not only rapidly cultivate students' ability to solve clinical problems but also save teaching costs, improve teaching quality and efficiency, and achieve the sharing of high-quality teaching resources. In addition to enriching the content of clinical teaching, students with diverse experience and innovative thinking are trained, which promotes the basic clinical knowledge of medical students and effectively cultivates their clinical thinking ability.

Acknowledgments

The authors would like to thank all participants for willingly sharing their experiences. This study was supported by grant #22-458 Development and implementation of an IT-based competency assessment system for Internal Medicine and Obstetrics & Gynecology resident training programs in China. The study funders had no rule in the study design, implementation, analysis, manuscript, preparation, or decision to submit this article for publication.

Data Availability

The datasets generated during and/or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

YW, NS, and JQ contributed equally to this study. YW and HW conceptualized the study and collected the data. HW and DZ jointly analyzed the data and drafted the article, which was later refined by YW, who also drafted and revised the paper. MW and HJ contributed to data analysis and offered comments. JH and NS provided critical comments and reviewed the important intellectual content. JQ contributed to conception and design as well as the final approval of the version to be published as the corresponding author.

Conflicts of Interest

None declared.

References

1. Richards JB, Hayes MM, Schwartzstein RM. Teaching clinical reasoning and critical thinking: from cognitive theory to practical application. Chest 2020 Oct;158(4):1617-1628. [doi: 10.1016/j.chest.2020.05.525] [Medline: 32450242]
2. Reinus JF. Patient-data management and clinical reasoning: neglected essential skills. Am J Med 2022 Dec;135(12):1393-1394. [doi: 10.1016/j.amjmed.2022.08.022] [Medline: 36063858]
3. Karunaratne D, Sibbald M, Chandratilake M. Understanding cultural dynamics shaping clinical reasoning skills: A dialogical exploration. Med Educ 2024 Aug 1. [doi: 10.1111/medu.15479] [Medline: 39089689]
4. Locke R, Mason A, Coles C, Lusznat RM, Masding MG. The development of clinical thinking in trainee physicians: the educator perspective. BMC Med Educ 2020 Jul 16;20(1):226. [doi: 10.1186/s12909-020-02138-w] [Medline: 32678045]
5. Lee DS, Abdullah KL, Subramanian P, Bachmann RT, Ong SL. An integrated review of the correlation between critical thinking ability and clinical decision-making in nursing. J Clin Nurs 2017 Dec;26(23-24):4065-4079. [doi: 10.1111/jocn.13901] [Medline: 28557238]
6. Carin L. On artificial intelligence and deep learning within medical education. Acad Med 2020 Nov;95(11S Association of American Medical Colleges Learn Serve Lead: Proceedings of the 59th Annual Research in Medical Education Presentations):S10-S11. [doi: 10.1097/ACM.0000000000003630] [Medline: 32769462]
7. Myers JH, Dorsey JK. Using diagnostic reasoning (DxR) to teach and evaluate clinical reasoning skills. Acad Med 1994 May;69(5):428-429. [doi: 10.1097/00001888-199405000-00052] [Medline: 8086079]
8. Xu X, Liu F. Optimization of online education and teaching evaluation system based on GA-BP neural network. Comput Intell Neurosci 2021;2021:8785127. [doi: 10.1155/2021/8785127] [Medline: 34422036]
9. Nagi F, Salih R, Alzubaidi M, et al. Applications of artificial intelligence (AI) in medical education: a scoping review. Stud Health Technol Inform 2023 Jun 29;305:648-651. [doi: 10.3233/SHTI230581] [Medline: 37387115]
10. Hsiao CT, Chou FC, Hsieh CC, Chang LC, Hsu CM. Developing a competency-based learning and assessment system for residency training: analysis study of user requirements and acceptance. J Med Internet Res 2020 Apr 14;22(4):e15655. [doi: 10.2196/15655] [Medline: 32286233]
11. Webster CS. Artificial intelligence and the adoption of new technology in medical education. Med Educ 2021 Jan;55(1):6-7. [doi: 10.1111/medu.14409] [Medline: 33155316]
12. Wang M, Sun Z, Jia M, et al. Intelligent virtual case learning system based on real medical records and natural language processing. BMC Med Inform Decis Mak 2022 Mar 4;22(1):60. [doi: 10.1186/s12911-022-01797-7] [Medline: 35246134]
13. Meng Q, Zhaoxia T. Promoting effect of information products on case-based clinical teaching model. J Med Inform 2012(8):89-92. [doi: 10.3969/j.issn.1673-6036.2012.08.024]
14. Han ER, Yeo S, Kim MJ, Lee YH, Park KH, Roh H. Medical education trends for future physicians in the era of advanced technology and artificial intelligence: an integrative review. BMC Med Educ 2019 Dec 11;19(1):460. [doi: 10.1186/s12909-019-1891-5] [Medline: 31829208]

Abbreviations

AI: artificial intelligence
CBLAS: competency-based learning and assessment system
DxR: diagnostic reasoning

Edited by A Mavragani; submitted 26.04.24; peer-reviewed by D Hu, Q Xia; revised version received 23.09.24; accepted 24.09.24; published 03.01.25.

Please cite as:
Wang H, Zheng D, Wang M, Ji H, Han J, Wang Y, Shen N, Qiao J
Artificial Intelligence–Powered Training Database for Clinical Thinking: App Development Study
JMIR Form Res 2025;9:e58426
URL: https://formative.jmir.org/2025/1/e58426
doi:10.2196/58426

Original Paper

The Physical Activity at Work (PAW) Program in Thai Office Workers: Mixed Methods Process Evaluation Study

Katika Akksilp[1,2], MD, PhD; Thomas Rouyard[3], PhD; Wanrudee Isaranuwatchai[2], PhD; Ryota Nakamura[4], PhD; Falk Müller-Riemenschneider[1,5,6], PhD; Yot Teerawattananon[1,2], PhD, MD; Cynthia Chen[1,5], PhD

[1]Health Intervention and Technology Assessment Programme, Ministry of Public Health, Nonthaburi, Thailand

[2]Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, Singapore

[3]Graduate School of Public Health & Health Policy, City University of New York, New York, NY, United States

[4]Hitotsubashi Institute for Advanced Study, Hitotsubashi University, Tokyo, Japan

[5]Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore, Singapore

[6]Digital Health Center, Berlin Institute of Health, Charité-Universitätsmedizin Berlin, Berlin, Germany

Corresponding Author:
Katika Akksilp, MD, PhD
Health Intervention and Technology Assessment Programme
Ministry of Public Health
6th Floor, 6th Building, Department of Health
Tiwanon Rd., Muang
Nonthaburi, 11000
Thailand
Phone: 66 2 590 4549
Fax: 66 2 590 4369
Email: gochi.akk@gmail.com

Abstract

Background: An increasing number of multicomponent workplace interventions are being developed to reduce sedentary time and promote physical activity among office workers. The Physical Activity at Work (PAW) trial was one of these interventions, but it yielded an inconclusive effect on sedentary time after 6 months, with a low uptake of movement breaks, the main intervention component.

Objective: This study investigates the factors contributing to the outcomes of the PAW cluster randomized trial.

Methods: Following the Medical Research Council's guidance for process evaluation of complex interventions, we used a mixed methods study design to evaluate the PAW study's recruitment and context (how job nature and cluster recruitment affected movement break participation), implementation (dose and fidelity), and mechanisms of impact (assessing how intervention components affected movement break participation and identifying the facilitators and barriers to participation in the movement breaks). Data from accelerometers, pedometers, questionnaires, on-site monitoring, and focus group discussions were used for the evaluation. Linear mixed effects models were used to analyze the effects of different intervention components on the movement breaks. Subsequently, qualitative analysis of the focus group discussions provided additional insights into the relationship between the intervention components.

Results: The participation in movement breaks declined after the third week, averaging 12.7 sessions (SD 4.94) per participant per week for the first 3 weeks, and continuing to decrease throughout the intervention. On-site monitoring confirmed high implementation fidelity. Analysis of Fitbit data revealed that each additional movement break was associated with a reduction of 6.20 (95% CI 6.99-5.41) minutes in sedentary time and an increase of 245 (95% CI 222-267) steps. Regarding the mechanisms of impact, clusters with higher baseline sedentary time demonstrated greater participation in movement breaks, while those with frequent out-of-office duties showed minimal engagement. Moreover, clusters with enthusiastic and encouraging movement break leaders were associated with a 24.1% (95% CI 8.88%-39.4%) increase in participation. Environmental and organizational support components using posters and leaders' messages were ineffective, showing no significant change in percentage participation in movement breaks (4.49%, 95% CI –0.49% to 9.47% and 1.82%, 95% CI –2.25% to 5.9%, respectively). Barriers such as high workloads and meetings further hindered participation, while the facilitators included participants' motivation to feel active and the perceived health benefits from movement breaks.

Conclusions: Despite high fidelity, the PAW trial did not significantly reduce sedentary time, with limited uptake of movement breaks due to context-related challenges, ineffective environmental support, and high workloads during the COVID-19 pandemic.

(JMIR Form Res 2025;9:e57604) doi:10.2196/57604

KEYWORDS

process evaluation; sedentary behavior; physical activity; workplace; movement breaks

Introduction

Background

Extensive research has examined workplace interventions to reduce sedentary behavior and promote physical activity globally [1,2]. Numerous interventions have yielded noteworthy results, showcasing their effectiveness, whereas some have encountered challenges in achieving desired outcomes [2-5]. While early meta-analyses, such as one from 1998, suggested minimal impact of workplace interventions on physical activity [6], subsequent reviews, including one from 2003, began to recognize their potential benefits [7]. However, more recent evidence has provided a clearer picture, indicating that more successful interventions tended to use pedometers, apply internet-based approaches, and include activities at social and environmental levels [5]. One systematic review found that physical environment interventions, whether implemented at the workstation, building, or neighborhood level, were effective in promoting physical activity in the workplace [8]. Finally, complex or multicomponent interventions emerge as the most effective workplace strategy for reducing sitting time [4], highlighting the intricate process leading to successful behavioral changes [9].

However, multicomponent intervention trial reports often fail to elucidate the rationale and mechanisms behind their results, leaving readers with questions regarding the efficacy of the interventions. Many scholars have criticized these trials as being akin to a “black box” because the underlying reasons for their success or failure remain unknown [10]. As a result, process evaluations, which assess how interventions were implemented in practice, are essential for deciphering the implications of the results from multicomponent intervention trials. These evaluations highlight aspects of the intervention that may require improvement to increase the probability of success [11,12].

Several frameworks advocate for process evaluations to explore the context, implementation, and impact mechanisms of intervention programs, encompassing aspects such as recruitment, reach, dose, fidelity, and challenges [10-13]. Process evaluations have become particularly relevant because they can help determine whether success or failure lies in the implementation, the intervention itself, or a combination of both factors. Notably, the Medical Research Council’s guidance on process evaluation of complex interventions is particularly valuable for assessing multicomponent interventions [11]. By contrast, frameworks like the one proposed by Steckler and Linnan [13] focus more on specific elements such as dose and fidelity but may not provide the broader perspective necessary to understand the interactions between context, mechanisms, and implementation. The emphasis of the Medical Research Council’s guidance on context and mechanisms was pivotal in capturing the socioecological dynamics within our intervention.

In addition, there is increasing acknowledgment that incorporating both qualitative and quantitative data, as well as using theoretical frameworks within process evaluations, plays a crucial role in promoting evidence-based practice [14]; for example, the Older People’s Exercise Intervention in Residential and Nursing Accommodation trial [15] used a mixed methods process evaluation to assess whether the multicomponent intervention changed residents’ home culture to increase physical activity and whether residents engaged in exercise activities. The study found no notable cultural shifts or sufficient engagement, thus explaining the trial’s null results [15,16]. Another process evaluation explored the implementation and impact mechanisms of a park prescription intervention trial, revealing key mediators of the intervention effects, such as park physical activity levels, as well as barriers that may have weakened intervention effectiveness [17].

Objectives

In Thailand, it has been reported that the majority of adults who engage in sufficient physical activity reside in rural areas, typically due to their work in the agricultural sector, whereas physical activity levels tend to be lower among office workers [18]. Numerous nonresearch initiatives have been launched to promote physical activity and reduce sedentary behavior among office workers in Thai companies and organizations. The Physical Activity at Work (PAW) trial [19] marked the first comprehensive cluster randomized trial aimed at promoting physical activity and reducing sedentary behavior among Thai office workers. This trial incorporated a multicomponent intervention and used accelerometer-measured data, laying a solid foundation for physical activity research in Thailand [20]. However, the trial produced inconclusive findings, leaving critical questions regarding the factors contributing to its outcomes unanswered. We observed a suboptimal uptake of movement breaks, the primary intervention component, and inferred that this, coupled with a low recruitment rate, may explain the absence of statistically significant outcomes [20]. Understanding these underlying reasons is essential for making necessary improvements in both research and policy. To address this, we conducted a process evaluation to investigate the factors contributing to the trial’s outcome.

Following the Medical Research Council’s guidance for process evaluation of complex interventions [11], we used a mixed methods study design to evaluate the PAW study’s recruitment and context, implementation, and impact mechanisms. Specifically, we examined how job nature and recruitment processes within each cluster influenced participation in movement breaks (context), the overall dose and fidelity of the

intervention (implementation), and the effects of intervention components on participation in movement breaks as well as the facilitators and barriers to movement break participation (impact mechanisms).

Methods

Description of the Cluster Randomized Trial

The PAW study was conducted in a free-living setting at the Ministry of Public Health offices in Nonthaburi, Thailand. Detailed methodological information regarding the cluster randomized trial (Thai Clinical Trials Registry: TCTR20200604007) was published in the protocol [19] and main results [20] manuscripts. Briefly, between July and September 2021, we recruited 282 office workers (age: mean 38.6, SD 10.4, years; female: n=228, 80.9%) from 18 separate offices. After baseline data collection, the participants were randomized into 9 intervention offices (140/282, 49.6%) and 9 control offices (142/282, 50.4%). The 6-month intervention, which took place between September 2021 and March 2022, consisted of 6 components across 4 levels, as outlined in Textbox 1.

Textbox 1. Key elements of the Physical Activity at Work cluster randomized trial.

Individual-level components

- Providing a wearable device with real-time activity feedback (Fitbit Inspire HR; Google LLC)
- Using a Fitbit smartphone app
- Offering individual weekly lottery-based financial incentives, where 1 intervention group participant who participated in at least 70% of the movement breaks within the previous week was randomly selected to receive a reward of THB 500 (US $16)

Social-level components

- Team movement breaks of light to moderate intensity, lasting at least 4 minutes and occurring 4 times a day (alarm clocks with speakers were provided to movement break leaders to initiate sessions), which served as the primary intervention component
- Team-based incentives of an additional weekly lottery reward of THB 500 (US $16) given to the winner if at least 70% of the participants in the cluster also attended at least 70% of the movement breaks within the previous week; 4 alarm reminders were set (9:30 AM, 10:30 AM, 2:30 PM, and 3:30 PM), and trained movement break leaders managed the starting times, songs, and movements; participants working from home were encouraged to join the sessions via web conferencing

Environmental-level component

- Three types of posters providing information on health risks associated with high sedentary time and the benefits of physical activity, as well as examples of stretching exercises

Organizational-level component

- Leadership support in the form of messages sent by office directors twice a week via Line (LY Corporation) to encourage participants to reduce sedentary time using movement breaks and increase physical activity, as well as to announce reward winners with congratulatory messages and photographs of the reward ceremony

The multicomponent intervention was developed using the socioecological model [9,21]. The intervention components were strategically designed to complement one another, with a particular focus on the primary component: the movement breaks. In addition, department directors actively encouraged participants to attend more movement break sessions and announced the weekly lottery reward, which was contingent upon the frequency of participation in the movement breaks. Two movement break leaders per office were trained to oversee these sessions. Moreover, posters were displayed to inform participants about the adverse effects of prolonged sedentary behavior and to provide examples of strategies to break up sedentary time in the office. All intervention participants were equipped with a Fitbit Inspire HR smartwatch (Google LLC), preset to remind them to interrupt prolonged sitting every 30 minutes. Consequently, the intervention operated as illustrated in Figure 1.

We formulated our research questions to encompass the context, implementation, and impact mechanisms, with a specific emphasis on the primary intervention component: the movement breaks. This emphasis stemmed from two key considerations: (1) as previously noted, our intervention design places the movement break as the central intervention component, supported by other components; and (2) we observed a low participation rate (median 31.5%, IQR 20.4%-42.7%) in the movement break sessions. Moreover, 77.5% (107/138) of the intervention participants attended fewer than half of the breaks (240/480, 50%) throughout the intervention period. We hypothesized that the absence of a significant intervention effect could be attributed to low attendance in the movement break sessions [20]. Hence, our focus is to investigate the underlying reasons behind this issue. Table 1 describes the process evaluation components, specific questions, and the data sources.

Figure 1. The mechanism of associations between the Physical Activity at Work cluster randomized trial intervention components and outcomes.

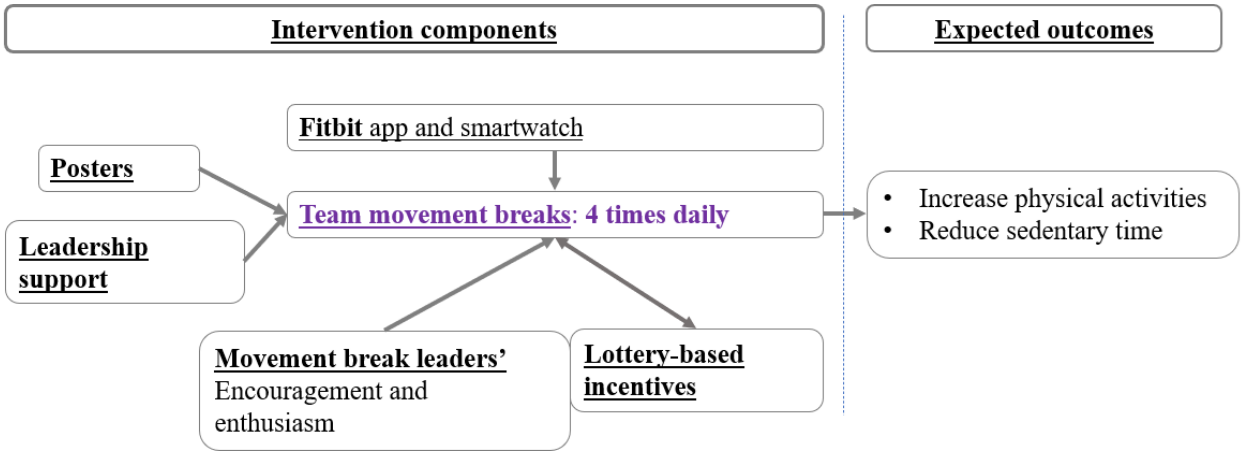


Table 1. Descriptions of the Physical Activity at Work process evaluation components, specific questions, and the data sources.

Process evaluation components	Specific process evaluation questions	Quantitative data	Qualitative data
Recruitment and context			
Recruitment	How does recruitment influence participation in movement breaks?	Recruitment rate	Recruitment fashion
Context	What impact do job descriptions and baseline characteristics of each cluster have on movement break participation?	Baseline summary statistics of each cluster	Job description of each cluster
Implementation			
Dose	What was the extent of intervention delivery?	Web monitoring: Fitbit wear time and conducts of movement breaks	On-site monitoring
Fidelity	Was the intervention delivered as intended?	Associations between participation in movement breaks and daily Fitbit sedentary time data as well as step counts during the intervention period	On-site monitoring
Mechanisms of impact			
Testing the assumptions behind the intervention design	How did supporting components influence participation in movement breaks?	Effects of intervention components on the percentage of weekly movement break participation	Focus group discussions
Participants' attitude	What were the facilitators and barriers to participation in movement breaks?	Additional questionnaire regarding facilitators and barriers to participation in movement breaks	Focus group discussions

Ethical Considerations

Ethics approval for both the quantitative and qualitative components of this study was granted by the Ethical Review Committee for Research in Human Subjects, Ministry of Public Health, Thailand (IRB00001629), in accordance with relevant ethical guidelines for human research. Any modifications to the approved protocol will be submitted to the ethics committee for review. All participants provided written informed consent before participation, which included consent for the use of deidentified data in secondary analyses and for the publication of results. To protect participants' privacy and confidentiality, all data were deidentified, and no personally identifiable information was included in the final dataset, with all published data presented in aggregate form. Participants' data (eg, case record forms, laboratory test results, information sheets, and consent forms) are stored in a locked cabinet in a researcher's office. All data will be destroyed by researchers within 5 years of publication. During the study, only deidentified data were used, and the data were only accessible to the research team. Participants received compensation of THB 250 (approximately US $7.5) for each data collection session in which they participated.

Dissemination Policy and Authorship Guidelines

In addition to disseminating our research findings to the funder of this study, the Ministry of Public Health, we will disseminate our findings to other countries, the study participants, and the research community. We have followed the authorship guidelines of the International Committee of Medical Journal Editing.

Data Collection

Baseline and Follow-Up Data Collection (Questionnaire)

An interviewer-administered questionnaire, based on the Thai National Statistical Office's health survey and capturing

sociodemographic data such as age and education [22], was used to collect participant data at both baseline and follow-up. In addition, intervention participants were asked supplementary questions regarding the implementation of the intervention during the follow-up data collection session (Multimedia Appendix 1).

Data Monitoring (Movement Break Schedule)

A data monitoring team, consisting of the trial implementer and an administrator, conducted weekly random field visits to 4 offices during the scheduled movement break sessions, adhering to the initial timetable. Monitored data included (1) the occurrence of sessions, (2) session quality, (3) participant attendance, and (4) any issues encountered during the session.

Movement break leaders were responsible for submitting weekly schedules via Microsoft Excel Online, indicating session timings and dates. Updates were required if a session did not fall within the 30-minute alarm windows (9:30 AM, 10:30 AM, 2:30 PM, and 3:30 PM), indicating absence or starting times outside the specified windows.

Fitbit Data and Movement Break Participation

During the 6-month intervention period, we collected Fitbit data, which included wear time, sedentary time, and step counts. Daily activity data were anonymized and used to compile weekly lists of potential reward recipients and to assess compliance with movement breaks. We determined participation in each movement break session by (1) analyzing a quadratic function of steps, (2) detecting quadratic functions within the 1-hour time frame of movement breaks, and (3) ensuring that participants had accumulated at least 100 steps.

Quantitative Data Analysis

Linear mixed models were used to examine the associations between intervention components and activity outcomes, as well as between the intervention components themselves. These models incorporated random intercepts from both individual and cluster levels, while also accounting for random slope by including intervention weeks. Weekly individual data were derived from daily Fitbit records, follow-up questionnaire responses, and monitoring data (Table 2). Data were analyzed using RStudio (version 4.0.3; Posit Software, PBC) and Stata software (version 14.2; StataCorp LLC), with significance level set at 5%.

Table 2. Variables used in the Physical Activity at Work process evaluation.

Variable	Explanation	Scale
Associations between movement break participation and daily Fitbit sedentary time and step count data during the intervention period		
Outcome variables		
Sedentary time	Daily sedentary time data from Fitbit	Continuous (min)
Steps	Daily step count data from Fitbit	Continuous (step count)
Exposure variables		
Movement break participation	Daily movement break participation	Ordinal; 1-4 (sessions)
Effects of intervention components on weekly movement break participation percentage		
Outcome variable		
Movement break participation percentage	The frequency of an individual's participation in movement break sessions within a specific week compared to the maximum number of sessions available during that week	Continuous (percentage)
Exposure variables		
Individual reward	We designated the weeks following individual reward wins as {1}, while all other weeks were coded as {0}, using the web monitoring data	1=won a reward last week; 0=others
Cluster reward	We assigned {1} to the weeks following office colleagues' reward wins and {0} to all other weeks, using the web monitoring data	1=a colleague in the same office cluster won a reward last week; 0=others
Fitbit wear time	We categorized the average weekly Fitbit wear time for each individual into a binary variable using the median	1=wore Fitbit at or above the median wear time; 0=wore Fitbit below the median wear time
Fitbit wear time (alternative)	How often Fitbit led individuals to engage in physical activity in the office in the last 2 weeks of the intervention	1=at least 4 d/wk; 0=<4 d/wk
Leadership support	How often the directors' support led individuals to engage in physical activity in the office in the last 2 weeks of the intervention	1=at least 4 d/wk; 0=<4 d/wk
Movement break leaders	How the movement break leaders' (1) encouragement and (2) enthusiasm contributed to movement break participation (We asked the movement break leaders and used the data as cluster-representative data)	3=both factors contributed *somewhat* or *a lot*; 2=only the encouragement contributed *somewhat* or *a lot*; 1=only the enthusiasm contributed *somewhat* or *a lot*; 0=both factors contributed *not at all*, *very little*, or *a little*
Posters	How often the posters led individuals to engage in physical activity in the office in the last 2 weeks of the intervention	1=at least 4 d/wk; 0=<4 d/wk
Posters (alternative)	How accurate individuals were in reporting the number of different poster designs in their office	1=3-4 designs; 0=0-2 designs

Focus Group Discussions

The qualitative aspect of the study involved focus group discussions aimed at exploring participants' perspectives on facilitators and barriers to engagement in movement breaks, as well as their attitudes toward intervention components. Intervention clusters were initially ranked by mean percentage participation in movement breaks. Subsequently, using purposive sampling, we invited up to 6 participants with the highest percentage participation from the top 2 clusters and up to 6 participants with the lowest percentage participation from the bottom 2 clusters for each of the 4 focus group discussions.

The 4 focus group discussions took place via 45- to 75-minute Zoom (Zoom Video Communications, Inc) meetings, with participants joining from their respective offices. The focus group discussions were recorded with participants' consent (both video and audio were captured), while notes were taken during the sessions. Before each focus group discussion, participants were briefed about its purpose, format, and estimated duration, with the option to interrupt as needed.

Verbatim transcriptions were manually completed by hired research assistants. The transcriptions were then subjected to deductive thematic analysis, with facilitator and barrier serving as overarching themes, and the socioecological model used as subtheme [9]. Two analysts (KA and WI) independently coded each transcript using the pre-established framework. References under themes and subthemes were compared and discussed to ensure consistency and accuracy. Intercoder reliability was not analyzed in this study.

An interview guide (Table S1 in Multimedia Appendix 2) was developed based on the research question "What are the facilitators and barriers to movement break participation?" The socioecological model informed this guide. A total of 4 interviewers (female: n=2, 50%) conducted the focus group discussions. They comprised the PAW trial implementer (first author; doctor of medicine), 2 study administrators (bachelor

of communication arts and master of political science), and a research assistant (bachelor of clinical pharmacology) who was not a trial staff member. All interviewers underwent a comprehensive 2-day training program tailored to the qualitative data collection and analysis requirements of the study. This training was led by a Thai senior researcher (PhD in anthropology). The focus group discussion guide served as a reference tool, ensuring that critical topics were not overlooked. During the focus group discussions, the researcher flexibly adjusted the sequence, content, and style based on individual responses. Emotions and nonverbal cues expressed by participants were carefully documented.

Code names were used in place of real names in the recorded data to protect participant privacy. Data collection and analysis proceeded concurrently throughout the study. After coding the transcripts from the 4 focus group discussions, the researchers and supervisors deliberated on data saturation. Subsequently, 2 additional focus group discussions were conducted with participants from clusters ranked in the top and bottom thirds, following the same procedures. No new themes emerged from these final 2 focus group discussions, indicating that data saturation had been achieved.

Results

Recruitment and Context

The PAW study recruited participants from all clusters within the Department of Medical Services building and the International Health Policy Program, Ministry of Public Health, Thailand, between July and September 2020. The ministry's governance structure is typically bureaucratic, with each department led by a single director general and each office headed by an office director. In total, 18 clusters were successfully recruited, with 15 (83%) falling under the Department of Medical Services purview; the remaining 3 (17%) clusters were under the jurisdiction of the Office of the Permanent Secretary (n=2, 67% were part of the International Health Policy Program and located in a distinct building complex; Figure S1 in Multimedia Appendix 2).

Recruitment procedures at the cluster level proceeded seamlessly. The directors of each office endorsed the active involvement of their respective teams in the trial, resulting in a cluster-level recruitment rate of 100% (18/18). For individual-level recruitment, we initiated the process by organizing office-specific group meetings to thoroughly explain the trial details. After these meetings, office workers interested in participating could engage with the trial staff to complete the individual informed consent process. To accommodate those unable to attend the scheduled meetings due to prior commitments, additional sessions were arranged.

Despite the bureaucratic governance style, the individual recruitment rate reached 62.9% (282/449). Job descriptions varied and included research-related roles and academic, finance, law, digital, and other administrative positions. At baseline, 3 (17%) of the 18 clusters exhibited a mean daily sedentary time exceeding 9 hours (mean sedentary time of 546, SD 90.3 min). Moreover, mean daily moderate to vigorous physical activity time was generally higher than the current physical activity guideline, reflecting that participants were relatively active. No significant associations were observed among job descriptions, baseline time spent in sedentary or moderate to vigorous physical activity, and monitoring data. However, clusters with higher baseline sedentary time demonstrated increased participation in movement breaks (Figure S2 in Multimedia Appendix 2). Furthermore, mean age and cluster size seemed to influence participation in movement breaks within the intervention group. Notably, individuals in the cluster with the youngest and smallest number of participants showed minimal participation in movement breaks (Table 3).

Table 3. Recruitment and context of intervention participants and the implementation of the Physical Activity at Work intervention (n=140).

	Baseline (n=140)			Data monitoring (6-mo intervention period) (n=138)			
Job description	Cluster size, n (%)	Recruitment rate n/N (%)	Sedentary time (min/d), mean (SD)	Lottery-incentive wins, n (%)	Movement break participation (%), mean (SD)	Maximum work from home, n/N (%)	Notes
Research related	23 (16.4	23/25 (92)	541 (90.5)	15 (63)	59.7 (14.7)	2.5/5 (50)	—[a]
Research related	8 (5.7)	8/44 (18)	483 (99.4)	0 (0)	25.8 (13.2)	5/5 (100)	Different building and department
Nursing	13 (9.3)	13/37 (35)	463 (117)	4 (17)	38.8 (16.6)	0/5 (0)	Different department
Finance	10 (7.1)	10/11 (90)	460 (95.9)	0 (0)	32.9 (5.49)	2/5 (40)	Broadcast from another cluster
Finance	14 (10.0)	14/17 (82)	481 (94.1)	1 (4.17)	29 (16.4)	2/5 (40)	—
Human resource	34 (24.3)	34/35 (97)	473 (113)	2 (8)	31.5 (11)	2.5/5 (50)	—
Human resource	18 (12.9)	18/31 (58)	394 (141)	1 (4)	25 (7.99)	2.5/5 (50)	Broadcast from another cluster
Digital	15 (10.7)	15/25 (60)	447 (116)	1 (4)	19.7 (9.83)	2/5 (40)	—
Inspection	5 (3.6)	5/18 (27)	350 (121)	0 (0)	3.88 (4.54)	2/5 (40)	n=3 (50%) participants at follow-up

[a]Not applicable.

Implementation

Fidelity: Was the Intervention Delivered as Intended?

On-Site Monitoring

The on-site monitoring team conducted 65 field visits to the intervention group participants' offices during their scheduled movement break sessions. During these visits, 35 movement breaks were observed, with 28 (80%) sessions featuring the Department of Medical Services theme song alongside another preferred Thai song. Importantly, there were no instances of cheating, such as merely shaking the Fitbit device without engaging in actual physical activity. However, 1 (8%) of the 13 participants from cluster 13 chose to take a 10-minute walk outside the office instead of participating in the team movement breaks, even when the sessions were not prompted. This participant managed to secure the weekly reward 3 times during the 24-week intervention period. Finally, no participants from the control group attended the sessions.

All alarm clocks were operational; nevertheless, leaders opted to broadcast the movement break songs using alternative devices, including the built-in broadcast system and office speakers. Notably, 2 (22%) of the 9 clusters did not independently initiate their movement break sessions; instead, they waited for another cluster to start the activity because they shared the same broadcast systems.

The movement break schedule reports submitted by the leaders were accurate; nonleaders who independently initiated sessions assisted with the reporting process. This collaboration ensured that the recorded data accurately reflected actual session initiation and participation. This also supports the finding that in some of the clusters (3/9, 33%), participants took the initiative to lead sessions without leader involvement, demonstrating shared responsibility in maintaining intervention fidelity.

Weekly rewards were distributed to winners by department directors. Participants received the team-based incentive on only 2 (8%) out of 24 occasions. A photograph capturing the reward ceremony, featuring face-to-face reward distribution, was shared with all intervention participants through Line OpenChat.

We received notifications regarding Fitbit issues, including syncing, freezing, and charging problems, both during and outside of our field visits. Fortunately, some participants with technological expertise were able to provide assistance, alleviating the need for the implementation team to respond to every notification. During our observations, we noticed that some of the participants (9/138, 6.5%) occasionally left their Fitbit devices at home when attending the movement break sessions. Despite understanding that their participation would not be recorded without the Fitbit, they still chose to engage in the sessions. In addition, 1 (0.7%) of the 138 participants was observed wearing the Fitbit only during the movement breaks.

All posters remained undamaged and visible, without any changes to their placement.

Due to the COVID-19 pandemic, a work-from-home policy was implemented. However, fewer than half of the participants (49/138, 35.5%) worked from home, with only 1 (11%) of the 9 clusters having a 100% work-from-home arrangement. This cluster used Zoom for meetings to conduct their movement breaks. Upon visiting the Zoom link twice, we found only 1 (0.7%) of the 138 participants in the Zoom meeting, with no movement breaks initiated. By contrast, another cluster coordinated their participation through a Line group chat when working from home, ensuring simultaneous session participation without requiring an online meeting.

Online Monitoring

We examined the correlations between engaging in movement breaks and daily Fitbit sedentary time and step count data. Each additional movement break was associated with a reduction of 6.20 (95% CI 6.99-5.41) minutes in sedentary time and an increase of 245 (95% CI 222-267) steps, adjusting for Fitbit wear time, number of public holidays in that week, and participants' age, sex, and education, with cluster and ID as random intercepts and intervention week number as a random slope.

This analysis, based on pedometer-measured outcomes, supported the fidelity of the movement break implementation because our prescribed minimum requirement for the duration of a single session was 4 minutes. Furthermore, the cumulative steps surpassed the eligibility criterion of 100 steps. The analysis also suggests successful data synthesis regarding movement break participation (Table 4).

Table 4. Associations between movement break participation on daily Fitbit sedentary time and step counts during the intervention period in the Physical Activity at Work cluster randomized trial.

	Model A		Model B	
	β[a] (95% CI)	*P* value	β[b] (95% CI)	*P* value
Sedentary time (min)	–6.32 (–7.06 to –5.57)	<.001	–6.20 (–6.99 to –5.41)	<.001
Steps (count)	263 (242 to 283)	<.001	245 (222 to 267)	<.001

[a]Linear mixed-effect model adjusted for Fitbit wear time, with cluster and ID as random intercepts and intervention week number as the random slope.

[b]Further adjusted for the number of public holidays in that week, age, gender, and education of the participants.

Dose: What Was the Extent of Intervention Delivery? (Online Monitoring)

According to the movement break leaders' weekly reports, many movement break sessions were never initiated by the leaders in each cluster. By the third week of the intervention, approximately 40% (72/180) of the scheduled movement break sessions had not been initiated (Figure S3 in Multimedia Appendix 2). Similarly, the overall average participation in movement breaks declined after the third week of the intervention, reducing to an average of 8 sessions per week per participant. Subsequently, the participation rate continued to decrease (Figure S4 in Multimedia Appendix 2).

We plotted the average daily Fitbit wear time to assess adherence during the intervention period (Figure S5 in Multimedia Appendix 2). We found that participants typically wore the devices for 10 to 15 hours per day at the start, with a slight decrease to 8 to 14 hours occurring 2 to 3 months into the intervention.

Mechanisms of Impact

Overview

Building on Figure 1, we present the effects of each component derived from a mixed methods analysis aimed at exploring the associations between each intervention component and movement break participation. We combine results from both quantitative and qualitative analyses to comprehensively address the question.

Quantitative Analysis

A total of 3200 participant-week data points were extracted by combining participants' demographics, Fitbit wear time, sedentary time and step count data, answers from the attitude-toward-intervention-components questionnaire, and online monitoring of rewards (Table 2).

Qualitative Analysis

Overview

There were 28 participants (n=3, 11% to n=6, 21% in each session) across 6 focus group discussions. The primary data analysts for the main results included 4 interviewers: the head of the data collection team (male), a program administrator (female), a researcher (male), and a staff member from the Health Intervention and Technology Assessment Program (a semiautonomous research unit under the Ministry of Public Health; female). Demographic details of all participants from the clusters (coded aby C12, C13, C5, C8, C16, and C17) involved in the focus group discussions are presented in Table S2 in Multimedia Appendix 2. Generally, the mean age of participants from the 3 top-ranked clusters was higher than that of participants from the bottom-ranked clusters. The majority of the participants (≥80%; 23/28, 82%) were female. Notably, the cluster with the lowest participation rate of 2.1% (10/480) comprised only 3 (2.2%) of the 138 participants at the 6-month follow-up (Table S2 in Multimedia Appendix 2).

How Did the Supporting Components Influence Participation in Movement Breaks?

Lottery-Based Incentives

Individuals who won the weekly lottery rewards the previous week demonstrated an 8.64% (95% CI 0.985%-16.3%) increase in movement break participation compared to other data points (Table 5). However, it is crucial to note the possibility of an overestimation when comparing data from winners against those from nonwinners. To address this, we conducted another subgroup analysis exclusively including winners, revealing a 5.1% (95% CI –3.44% to –13.6%) increase in movement break participation. However, this increase lacked statistical significance due to the small sample size (Table S3 in Multimedia Appendix 2).

Table 5. Effects of the Physical Activity at Work intervention components on weekly movement break participation percentage.

	β[a] (%; 95% CI)	Standardized β (%; 95% CI)	*P* value
Individual reward	8.64 (.985 to 16.3)	.0257 (.00293 to .0484)	.03
Cluster reward	–.325 (–2.64 to 1.99)	–.004 (–.0325 to .0245)	.78
Fitbit wear time	3.96 (2.28 to 5.65)	.0696 (.04 to .0993)	.001
Leadership support	1.82 (–2.25 to 5.9)	.0309 (–.0381 to .1)	.38
Enthusiastic	2 (–17.2 to 21.2)	0.077 (–.663 to .817)	.84
Encouraging	1.49 (–18.8 to 21.8)	0.0576 (–.724 to .84)	.89
Both enthusiastic and encouraging	24.1 (8.96 to 39.2)	0.929 (.346 to 1.51)	.002
Posters	4.49 (–.493 to 9.47)	0.0695 (–.00764 to .147)	.08

[a]Linear mixed effects model adjusted for the previous-week movement break participation percentage, number of public holidays in that week, age, sex, and education of the participants, with cluster and ID as random intercepts and intervention week number as the random slope.

During the focus group discussions, some of the participants suggested that rather than a single winner receiving a substantial cash prize, there should be multiple winners receiving more affordable rewards:

> *Rather than cash, it should be acknowledgments, such as showing who reach this many steps so we can compete with each other.* [C13]

> *[C]ould be something cheaper, don't have to be cash. Cheap shirts will do!* [C10]

> *[I]t was indeed a motivating factor, but the conditions for obtaining it should be somewhat more lenient. That is, to make it accessible to a broader audience, but the current conditions seem to be quite high. If it were reduced a bit, say to 100, I believe that providing rewards would be motivating enough.* [C8]

A participant in the worst-performing cluster did not remember the details of the financial incentive:

> *Sorry but how much was the reward?* [C17]

Team-Based Incentives

We observed no difference in movement break participation percentage (–0.325%, 95% CI –2.64% to 1.99%) in the weeks after a colleague from the same office won lottery rewards compared to other weeks. Contrary to our assumptions during the intervention development phase, peer pressure seems to have had no discernible effects on motivating movement break participation, as indicated by the quantitative analysis model (Table 5).

Different ideas emerged during the focus group discussions, as outlined in the following paragraphs.

Peer support was mentioned among members in the best-performing cluster as an important motivator:

> *Because it helped the team. If we dance, someone gets 1000THB, if we don't, it's 500THB only.* [C12]

However, others expressed discouragement:

> *When I see others got it and I never won for weeks, I was disheartened.* [C12]

Some of the participants viewed rewards as not motivating:

> *I joined the sessions because I want to be healthier. Rewards are meaningless for me.* [C13]

Fitbit Wear Time

There was a 3.96% (95% CI 2.28%-5.65%) increase in movement break participation among individuals with higher Fitbit wear time compared to those with lower wear time, using the median weekly wear time as the cutoff (Table 5). However, this increase may be attributed to motivation triggered by Fitbit notifications (eg, reminders to break prolonged sitting) or potentially to an information bias, wherein participants who wore their Fitbit devices for longer periods were detected more frequently in movement break participation.

To address potential bias, we compared the result with another model where Fitbit exposure was based on self-report, using the question "How often did you look at the Fitbit tracker during the last 2 weeks of the intervention period?" The analysis revealed that looking at the tracker more frequently (≥4 times/wk) was associated with a 1.97% (95% CI –1.42% to 5.36%) higher movement break participation percentage, without statistical significance (Table S4 in Multimedia Appendix 2).

Qualitative analysis indicated that participants perceived Fitbit as beneficial not necessarily for encouraging additional movement breaks but rather for motivating increased exercise and providing real-time feedback:

> *I joined short breaks a lot at first, but after a while, I forgot...However, I always wear the watch; look at the daily data.* [C13]
>
> *I set my goal to walk 10,000 steps a day after I got the watch, and I succeeded!* [C16]
>
> *[S]aying Fitbit encourages more movement breaks is wrong for me, but I feel the urge to exercise more from wearing the watch, like running after work.* [C17]

Leadership Support

Participants perceived the encouragement from directors as ineffective in motivating them to participate in more movement breaks.

Although office directors joined very few movement breaks, participants believed that their presence helped motivate everyone in the cluster:

> *He joined, I saw. But mostly he's busy.* [C16]
>
> *It made us stand up and dance...if we didn't it'd be awkward.* [C5]
>
> *We all danced every time our director was present (laugh).* [C8]

With regard to the movement break leaders' encouragement and enthusiasm, clusters with leaders who self-evaluated as more encouraging and enthusiastic had a significant increase of 24.1% (95% CI 8.88%-39.4%) in movement break participation compared to clusters with less encouraging and less enthusiastic leaders. However, because these exposure variables are self-reported by cluster leaders, the detected difference may be influenced by other cluster-specific factors.

Participants believed that movement break leaders' enthusiasm and encouragement helped them join more break sessions:

> *He was very active and always encourages everyone to stand up and dance.* [C8]
>
> *We were aware of the scheduled time, but if we were occupied, the leader would notify us.* [C13]

Nevertheless, some of the participants mentioned that they did not rely on leaders:

> *No matter how many people in the office, we danced. no leaders, no problem at all.* [C12]
>
> *I'm not sure...leaders always initiated the activity, but we did not always join.* [C16]
>
> *No one was available.* [C17]

Posters

Participants who reported that posters motivated them to engage in more movement breaks during the last 2 weeks of the intervention exhibited approximately a 5% higher participation rate, although this increase lacked statistical significance (Table 5). Notably, this increase may indicate participants' attitudes toward the intervention component rather than the direct effects of the posters themselves. Therefore, we compared the result by converting the variable to whether participants accurately identified the number of different styles of posters in their offices. The analysis revealed no significant difference in

movement break adherence between those who answered correctly and those who did not (standardized β=1.94, 95% CI –1.34 to 5.23; Table S5 in Multimedia Appendix 2).

Qualitative analysis showed that while the posters initially captured interest, over time, they failed to sustain attention:

> *At first I read them. I thought it was helpful and tried to follow some moves. After a while I just ignored them.* [C2]

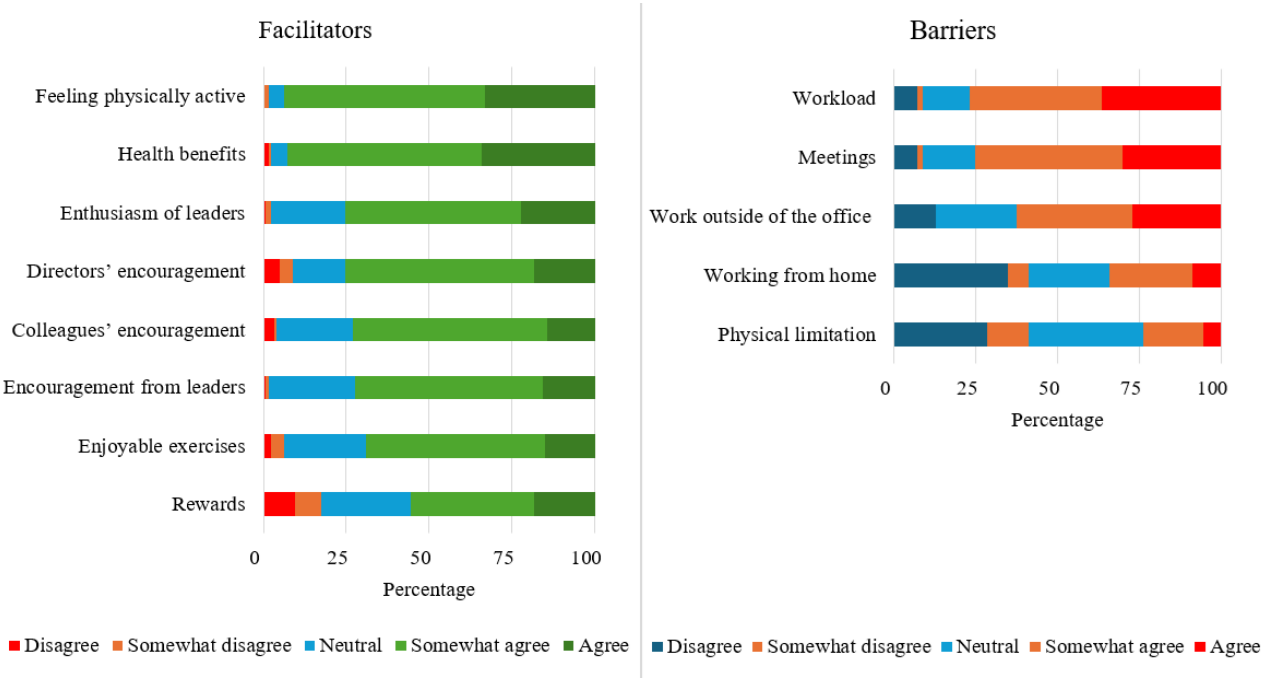
> *I didn't really read it that much...just walk past.* [C3]

The standardized β coefficients of all exposures indicate that the self-evaluated enthusiasm and encouragement of movement break leaders were the most important variables among those included, followed by Fitbit wear time and the individual reward (Table 5).

Facilitators and Barriers to Movement Break Participation

Figure 2 illustrates participants' attitudes regarding the facilitators and barriers to movement break participation. The top 2 facilitators were the positive feelings associated with being physically active and the perceived health benefits, followed by encouragement from leaders, directors, and colleagues. The appeal of engaging in enjoyable exercises also motivated their participation. Interestingly, weekly rewards ranked lowest among the facilitators.

Figure 2. Facilitators and barriers to movement break participation in the Physical Activity at Work cluster randomized trial.



Conversely, the primary barrier perceived by participants was their workload, followed by meetings and working outside of the office. Surprisingly, many participants did not view working from home as a barrier to their participation.

Thematic analysis revealed positive attitudes toward intervention components:

> *I think having Fitbit is a really good motivation to move.* [C17]

> *Previously when someone invited me to run, I wanted to but it was hard. This [movement breaks] is easier and fun.* [C12]

Participants also noted barriers to the intervention, such as the lottery reward design:

> *It [individual reward] somehow motivated me, but I guess I just could not do as good as others.* [C16]

Another important barrier was the monotonous design of the movement breaks:

> *Dancing with the same moves gets boring after a while.* [C8]

> *I think we can change the songs to make it more interesting.* [C16]

Table 6 presents thematic analysis results with references under each theme and subtheme from all 6 focus group discussions.

Table 6. Thematic analysis of focus group discussions on the facilitators and barriers to movement break participation in the Physical Activity at Work cluster randomized trial.

Themes and subthemes	Quotes
Facilitators	
Psychological capability	• "Previously when someone invited me to run, I wanted to, but it was hard. This is easier and fun." (C12)
Social opportunity	• "Both songs and moves...it was like mass hysteria; others enjoyed the sessions and I wanted to join" (C12) • "Our director joined, and we danced happily" (C8) • "When my boss dance and I just sat there working, it felt strange, so I joined" (C5)
Physical opportunity	• "I'm happy that there is a new opportunity for me to exercise...in workplace. Normally I don't have time." (C12) • "Our office is spacious. There are empty spaces for dancing." (C5)
Reflective motivation	• "My personal goal was actually health, not rewards" (C13) • "I set my goal to get the reward and lose weight and cholesterol level because the programme measured those" (C5)
Automatic Motivation	• "It was relaxing, both mind and body...especially the mind." (C17) • "Just that we got to dance...when the session started, it relieved the stress quite a lot" (C5)
Barriers	
Psychological capability	• "10 min. dancing is too short. 30 min. working out is better for your health for the day." (C13) • "My programming work never allows me to lose focus" (C16)
Social opportunity	• "Meetings definitely prevented us to join movement sessions" (C8) • "We were not really motivated by directors and colleagues" (C17)
Physical opportunity	• "Sometimes workspace is not wide enough...also I was afraid I'll annoy others who were not in the project" (C13) • "During high workload, we cannot join" (C5) • "The high workload never allows us to do anything else" (C17) • "Ever since COVID situation got worse, I've been sitting all the time at my desk. It's the workload, can't do anything else." (C16)
Reflective motivation	• "Rewards could be anything, like shoes, incentives don't need to be money, like shirts..." (C5)
Automatic motivation	• "Each programme shouldn't last long. I mean, we should always change the stimulant to avoid boredom..." (C13) • "Dancing with the same moves gets boring after a while" (C8) • "I think we can change the songs to make it more interesting." (C16)

Discussion

The PAW cluster randomized trial of a multicomponent intervention, developed based on the socioecological framework [9], showed a decrease in participants' waking sedentary time and an increase in moderate to vigorous physical activity, although without statistical significance [20]. We conducted a mixed methods process evaluation, following the Medical Research Council's guidance [11], to comprehensively describe the PAW trial's (1) recruitment and context, (2) implementation, and (3) impact mechanisms.

Recruitment and Context

Our team anticipated a high recruitment rate once the office directors approved the inclusion of their offices in the trial, given the bureaucratic nature of the ministry organization in Thailand. However, recruitment rates were low in many clusters, weakening the trial's statistical power and risking reduced intervention adherence, particularly in the social-level component.

Cluster 17 serves as a notable example, demonstrating minimal engagement in movement breaks (Table 3; Figure S4 in Multimedia Appendix 2). This could be attributed to the recruitment of too few, relatively young participants alongside a high number of nonparticipants (Tables 3 and 4). Despite its potential for easy qualification for the team-based incentive with just 3 participants (at follow-up), cluster 17 might have faced challenges due to the negative influence of an environment where breaks are discouraged. This aligns with a previous study reporting reduced break taking in disapproving work environments compared to more supportive ones [23]. Cluster 13 also faced challenges with a low recruitment rate; yet, the engagement of the participants in this cluster in movement breaks was high. Two reasons might explain this: (1) the participant who chose to walk outside the office and secured 3 weekly rewards increased the cluster's average participation

rate, and (2) recruiting at least 13 participants could create an environment conducive to making movement breaks feel enjoyable and secure to participate in.

Job characteristics might also play a role in movement break participation. Cluster 17, "the inspection office," frequently required its members to leave the office for inspections, which was reflected in their low sedentary time at baseline (Table 3) and echoed in the focus group interviews. In addition, cluster 16 represented a digital office where most participants worked on their laptop computers all day. While this setting might seem ideal for implementing the intervention, participants encountered challenges with breaks because they needed continual focus to code. This aligns with the suggestion that identifying the optimal fit between the organizational context and the intervention is crucial [24]. However, challenges remain when implementing such interventions across organizations with diverse offices, each characterized by unique job contexts and work styles.

Implementation

We implemented the trial with a comprehensive dose delivery, ensuring intervention fidelity by closely adhering to the original plan for each component. Moreover, the findings in Table 3 affirm the application of precise criteria for enumerating movement breaks, underscoring their correlations with the expected reductions in sedentary time and increases in step counts. Nevertheless, we encountered challenges related to the frequency of movement break sessions conducted by participants, which we refer to as "dose received" [13]. This inconsistency may be attributed to the performance of the movement break leaders, as illustrated in Figure S3 in Multimedia Appendix 2, which shows that many sessions were never initiated. However, several clusters relied on automatic timing mechanisms to initiate breaks, ensuring that sessions commenced even without leaders present. We discuss this further in the next subsection.

Mechanisms of Impact

Testing the Designed Intervention Components

Healthy behaviors are maximized when environments and policies support healthful choices, and individuals are motivated and educated to make them [25]. As one of the main principles of the ecological model, the interaction of influences means that variables within the system work together synergistically [26]. Hence, our study examined the impact of the supporting components on the movement break participation using multivariate linear mixed models (Table 4).

The enthusiasm and encouragement of the movement break leaders seem to have contributed to higher participation percentages. However, the cluster that exhibited the best performance challenged this observation, asserting that the members initiated sessions independently, even without leaders. Unlike other cluster randomized trials where exercise sessions were led by nonparticipants such as physiotherapists [16], the movement break leaders in our study were participants and could be replaced if absent. Hence, without the leaders, other members could initiate sessions automatically. Nevertheless, the influence of the leaders' encouragement and enthusiasm likely played a role in motivating the rest of the team. Prior research has also found that workplace team leaders play a significant role in facilitating the implementation of workplace interventions [27,28]. Moreover, a comprehensive review emphasizes that effective team performance underscores the fulfillment of leadership styles, supportive team behaviors, communication, and performance feedback [29].

The impact of the team-based incentive on movement break participation was evidently negligible. Regarding individual lottery-based incentives, their influence remains somewhat unclear. Systematic reviews suggest that financial rewards for physical activity have positive short-term effects, surpassing unconditional incentives [30,31]. Moreover, another study indicates that increasing reward values may lead to improved results [32]. Nevertheless, insights from online monitoring data and focus group discussions revealed that only 1 (11%) of the 9 clusters actively pursued rewards and engaged in friendly competition, while the others (8/9, 89%) were indifferent. Awarding only 1 winner per week might have demotivated participants who faithfully adhered to the intervention but never won. This phenomenon can be explained using goal-setting theory, which suggests that setting clear and appropriately challenging goals is crucial [33]. By contrast, setting unachievable goals may induce stress, anxiety, and perceived pressure [34]. Therefore, overly difficult goals, such as securing the weekly lottery reward, may have discouraged participants.

Fitbit was perceived as a helpful tool for real-time data monitoring and might help motivate leisure physical activity. However, current evidence indicates unfavorable outcomes regarding the effectiveness of pedometers in increasing physical activity within the workplace or in motivating sedentary breaks [35,36]. The component may have primarily served as a data collection tool rather than actively supporting movement breaks. By contrast, the posters, which served as the environmental component, proved ineffective and should be replaced with physical environment interventions, which have been shown to be more effective [4,8].

Finally, leadership support was considered helpful at the start of the intervention but provided no lasting effect over time. The component could be perceived as a nudge, which has been widely used to prompt behavioral change among participants by providing alternative options to sedentary habits in the workplace [37]. Leadership support was expected to facilitate these behavioral shifts. Management support has also been found to be an essential enabler for workplace intervention participation to reduce sitting time [27,38,39]. However, the sustainability of the effects depends on the meticulous design of the intervention component. Future studies evaluating long-term effects must enhance the intervention component to ensure sustainability.

Facilitators and Barriers

Automatic motivation, driven by the desire to feel active and relaxed, was ranked as the top facilitator of movement break participation, followed by reflective motivation, including perceived health benefits. The findings are in line with a recent systematic review [39]. Another review also supports the idea that microbreaks at work can improve well-being by boosting vigor and reducing fatigue [40]. In addition, participants

generally grasped the concept of health benefits associated with movement breaks. However, some of the individuals compared movement breaks to aerobic exercise, expressing doubt regarding their health benefits.

Workload and meetings were perceived as the main barriers to movement breaks, which aligns with a previous study's prediction that high workloads, although positively related to the desire to detach from work, would also deter employees from actually taking breaks [23]. In addition, recent findings from a systematic review supported the notion that microbreaks increase well-being but do not necessarily improve work performance, especially in tasks that require high cognitive engagement. Moreover, the review suggested that breaks of >10 minutes may be necessary to enhance work performance [40].

By contrast, meetings presented clear barriers to our intervention design. Future research should be dedicated to advocating feasible and context-specific active meetings within an active workplace; for example, an exploratory study found that standing meetings were feasible, well-received by employees, and may reduce sitting time among the population [41]. However, widespread adoption faces obstacles due to prevailing sedentary work cultures and concerns about self-consciousness in front of senior staff, highlighting the need for broader social behavior change initiatives [42].

We hypothesized that movement break participation might be low due to the work arrangement during the peaks of the COVID-19 pandemic [20]. However, participants did not consider working from home to be the main barrier to movement break participation. Instead, they thought that the COVID-19 pandemic did not significantly hinder their adherence to the intervention due to the work-from-home policy. Instead, they attributed the difficulty to increased workload. Nevertheless, working from home remained 1 of the barriers to movement break participation. This is in line with another workplace cluster randomized trial, which reported that engaging in physical exercise with colleagues during working hours was more effective than home-based exercise in enhancing vitality and managing pain-related concerns among health care workers [43].

Strengths and Limitations

We conducted a comprehensive mixed methods process evaluation of the PAW multicomponent intervention, covering recruitment and context, implementation, and impact mechanisms. Rigorous analyses were applied to both quantitative and qualitative data. The data are prospective, spanning 6 months of follow-up. Although the overall results show no significant impact of the intervention, this process evaluation offers insights that could be crucial for the future development and evaluation of intervention packages aimed at reducing sedentary time while improving physical activity levels.

Nevertheless, our study had some limitations. First, we could not test mediators to understand the underlying mechanisms through which one variable influences another due to the trial's lack of efficacy [20]. Second, constructing intervention theories will involve reconstructing a logic model [11,44], a task we plan to undertake in future studies. Third, we could not complete the management and analysis of process evaluation data before the conclusion of the PAW trial. As a result, the process evaluation was conducted post hoc to elucidate the trial outcomes. Furthermore, our evaluation only incorporated participants' data and perspectives, overlooking input from other stakeholders such as organizational directors and nonparticipants. Finally, the team members responsible for process evaluation were also involved in the outcome evaluation. While our team members possess the most comprehensive understanding of the trial details, potential bias in interpretations must be acknowledged [11].

Conclusions

The PAW trial did not significantly reduce sedentary time among Thai office workers. Although the trial implementation was satisfactory regarding dose delivery and fidelity, there was limited uptake of the movement breaks, the key intervention component. This limited uptake could be attributed to (1) context-related challenges (including jobs requiring high cognitive engagement or frequent out-of-office work and meetings), (2) the absence of goal-setting aspects in the detailed design of individual and social components, (3) the lack of effective and sustainable supporting components at the environmental and organizational levels, and (4) elevated workloads in specific clusters (exacerbated during peak periods of the COVID-19 pandemic) serving as a significant barrier.

Acknowledgments

The authors thank Buppa Pardang, Pimonphan Thongoun, Saudamini Dabak, Kewalin Chomrenoo, Nachawish Kittibovorndit, and Budsadee Soboon for their help with the focus group discussions, further helpful discussions, and administrative support that contributed to this work. The trial was funded by revenue from taxes on tobacco and alcohol via the Thai Health Promotion Foundation. This study was supported by the Singapore Ministry of Health's National Medical Research Council under its Population Health Research Grant (PHRGOC22Jul-0020, ID: MOH-001315). The Health Intervention and Technology Assessment Program's International Unit is supported by the International Decision Support Initiative to provide technical assistance on health intervention and technology assessment to governments in low- and middle-income countries. RN and TR are supported by the Japan Society for the Promotion of Science Core-to-Core Program (JPJSCCB20200002). The funders had no role in the study design, data collection or analysis, the preparation of the manuscript, or the decision to publish. The views and opinions expressed in this paper are those of the authors and do not necessarily reflect those of the funders or the institutions to which the authors belong.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

All authors contributed to the study design or delivery of the trial. KA was the principal investigator of the trial. FM-R and CC provided statistical expertise and data management. TR, RN, YT, WI, and FM-R provided expertise on conceptualizing the process evaluation aspects and the narrative of the manuscript. RN and TR provided expertise on behavioral economics, such as the impact of lottery-based and team-based incentives. TR and FM-R provided expertise on physical activity promotion theories and frameworks. KA, YT, and CC drafted the manuscript. All authors reviewed the manuscript draft and read and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional questions for intervention participants.
[DOCX File , 740 KB - formative_v9i1e57604_app1.docx]

Multimedia Appendix 2
Supplementary figures and tables.
[DOCX File , 15 KB - formative_v9i1e57604_app2.docx]

References

1. Buckingham SA, Williams AJ, Morrissey K, Price L, Harrison J. Mobile health interventions to promote physical activity and reduce sedentary behaviour in the workplace: a systematic review. Digit Health 2019 Mar 27;5:2055207619839883 [FREE Full text] [doi: 10.1177/2055207619839883] [Medline: 30944728]
2. Zhou L, Deng X, Xu M, Wu Y, Shang X, E F, et al. The effects of active workstations on reducing work-specific sedentary time in office workers: a network meta-analysis of 23 randomized controlled trials. Int J Behav Nutr Phys Act 2023 Jul 27;20(1):92 [FREE Full text] [doi: 10.1186/s12966-023-01467-5] [Medline: 37501138]
3. Mulchandani R, Chandrasekaran AM, Shivashankar R, Kondal D, Agrawal A, Panniyammakal J, et al. Effect of workplace physical activity interventions on the cardio-metabolic health of working adults: systematic review and meta-analysis. Int J Behav Nutr Phys Act 2019 Dec 19;16(1):134 [FREE Full text] [doi: 10.1186/s12966-019-0896-0] [Medline: 31856826]
4. Chu AH, Ng SH, Tan CS, Win AM, Koh D, Müller-Riemenschneider F. A systematic review and meta-analysis of workplace intervention strategies to reduce sedentary time in white-collar workers. Obes Rev 2016 May;17(5):467-481. [doi: 10.1111/obr.12388] [Medline: 26990220]
5. To QG, Chen TT, Magnussen CG, To KG. Workplace physical activity interventions: a systematic review. Am J Health Promot 2013;27(6):e113-e123. [doi: 10.4278/ajhp.120425-LIT-222] [Medline: 23631453]
6. Dishman RK, Oldenburg B, O'Neal H, Shephard RJ. Worksite physical activity interventions. Am J Prev Med 1998 Nov;15(4):344-361. [doi: 10.1016/s0749-3797(98)00077-4] [Medline: 9838977]
7. Proper KI, Koning M, van der Beek AJ, Hildebrandt VH, Bosscher RJ, van Mechelen W. The effectiveness of worksite physical activity programs on physical activity, physical fitness, and health. Clin J Sport Med 2003 Mar;13(2):106-117. [doi: 10.1097/00042752-200303000-00008] [Medline: 12629429]
8. Zhu X, Yoshikawa A, Qiu L, Lu Z, Lee C, Ory M. Healthy workplaces, active employees: a systematic literature review on impacts of workplace environments on employees' physical activity and sedentary behavior. Built Environ 2020 Jan;168:106455. [doi: 10.1016/j.buildenv.2019.106455]
9. Sallis J, Owen N, Fisher EB. Ecological models of health behavior. In: Glanz BK, K V, editors. Health Behavior and Health Education: Theory, Research, and Practice. Hoboken, NJ: Jossey-Bass; 2008.
10. Grant A, Treweek S, Dreischulte T, Foy R, Guthrie B. Process evaluations for cluster-randomised trials of complex interventions: a proposed framework for design and reporting. Trials 2013 Jan 12;14:15 [FREE Full text] [doi: 10.1186/1745-6215-14-15] [Medline: 23311722]
11. 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]
12. Oakley A, Strange V, Bonell C, Allen E, Stephenson J, RIPPLE Study Team. Process evaluation in randomised controlled trials of complex interventions. BMJ 2006 Feb 18;332(7538):413-416 [FREE Full text] [doi: 10.1136/bmj.332.7538.413] [Medline: 16484270]

13. Steckler A, Linnan L. Process Evaluation for Public Health Interventions and Research. San Francisco, CA: Jossey-Bass; 2022.
14. Liu H, Andersson G, Manchaiah V. Editorial: the process evaluation of clinical trials. Front Med (Lausanne) 2022 Jul 11;9:950637 [FREE Full text] [doi: 10.3389/fmed.2022.950637] [Medline: 35899205]
15. Ellard DR, Taylor SJ, Parsons S, Thorogood M. The OPERA trial: a protocol for the process evaluation of a randomised trial of an exercise intervention for older people in residential and nursing accommodation. Trials 2011 Feb 02;12:28 [FREE Full text] [doi: 10.1186/1745-6215-12-28] [Medline: 21288341]
16. Ellard DR, Thorogood M, Underwood M, Seale C, Taylor SJ. Whole home exercise intervention for depression in older care home residents (the OPERA study): a process evaluation. BMC Med 2014 Jan 03;12:1 [FREE Full text] [doi: 10.1186/1741-7015-12-1] [Medline: 24386900]
17. Petrunoff N, Yao J, Sia A, Ng A, Ramiah A, Wong M, et al. Activity in nature mediates a park prescription intervention's effects on physical activity, park use and quality of life: a mixed-methods process evaluation. BMC Public Health 2021 Jan 22;21(1):204 [FREE Full text] [doi: 10.1186/s12889-021-10177-1] [Medline: 33482787]
18. Tuangratananon T. Differences in physical activity levels between urban and rural adults in Thailand: findings from the 2015 National Health and Welfare Survey. Health Syst Res 2018;12(2018):27-41 [FREE Full text]
19. Chen C, Dieterich AV, Koh JJ, Akksilp K, Tong EH, Budtarad N, et al. The physical activity at work (PAW) study protocol: a cluster randomised trial of a multicomponent short-break intervention to reduce sitting time and increase physical activity among office workers in Thailand. BMC Public Health 2020 Sep 01;20(1):1332 [FREE Full text] [doi: 10.1186/s12889-020-09427-5] [Medline: 32873258]
20. Akksilp K, Koh JJ, Tan V, Tong EH, Budtarad N, Xueying G, et al. The physical activity at work (PAW) study: a cluster randomised trial of a multicomponent short-break intervention to reduce sitting time and increase physical activity among office workers in Thailand. Lancet Reg Health Southeast Asia 2022 Oct 19;8:100086 [FREE Full text] [doi: 10.1016/j.lansea.2022.100086] [Medline: 37384135]
21. Owen N, Sugiyama T, Eakin EE, Gardiner PA, Tremblay MS, Sallis JF. Adults' sedentary behavior determinants and interventions. Am J Prev Med 2011 Aug;41(2):189-196. [doi: 10.1016/j.amepre.2011.05.013] [Medline: 21767727]
22. The Health and Welfare Survey 2015. National Statistical Office of Thailand. 2023. URL: https://www.nso.go.th/nsoweb/storage/survey_detail/2023/20230505165728_20473.pdf [accessed 2024-12-17]
23. Phan V, Beck JW. Why do people (not) take breaks? An investigation of individuals' reasons for taking and for not taking breaks at work. J Bus Psychol 2023;38(2):259-282 [FREE Full text] [doi: 10.1007/s10869-022-09866-4] [Medline: 36593864]
24. Nielsen K, Taris TW, Cox T. The future of organizational interventions: addressing the challenges of today's organizations. Work Stress 2010 Oct 20;24(3):219-233. [doi: 10.1080/02678373.2010.519176]
25. Ottawa charter for health promotion. World Health Organization. 2012 Jun 16. URL: https://www.who.int/publications/i/item/WH-1987 [accessed 2024-10-31]
26. Glanz K, Viswanath K, Rimer BK. Health Behavior and Health Education: Theory, Research, and Practice. Hoboken, NJ: John Wiley & Sons; 2008.
27. Goode AD, Hadgraft NT, Neuhaus M, Healy GN. Perceptions of an online 'train-the-champion' approach to increase workplace movement. Health Promot Int 2019 Dec 01;34(6):1179-1190. [doi: 10.1093/heapro/day092] [Medline: 30452649]
28. Hopkins JM, Glenn BA, Cole BL, McCarthy W, Yancey A. Implementing organizational physical activity and healthy eating strategies on paid time: process evaluation of the UCLA WORKING pilot study. Health Educ Res 2012 Jun;27(3):385-398 [FREE Full text] [doi: 10.1093/her/cys010] [Medline: 22323279]
29. Salcinovic B, Drew M, Dijkstra P, Waddington G, Serpell BG. Factors influencing team performance: what can support teams in high-performance sport learn from other industries? A systematic scoping review. Sports Med Open 2022 Feb 22;8(1):25 [FREE Full text] [doi: 10.1186/s40798-021-00406-7] [Medline: 35192078]
30. Barte JC, Wendel-Vos GC. A systematic review of financial incentives for physical activity: the effects on physical activity and related outcomes. Behav Med 2017;43(2):79-90. [doi: 10.1080/08964289.2015.1074880] [Medline: 26431076]
31. Mitchell MS, Goodman JM, Alter DA, John LK, Oh PI, Pakosh MT, et al. Financial incentives for exercise adherence in adults: systematic review and meta-analysis. Am J Prev Med 2013 Nov;45(5):658-667. [doi: 10.1016/j.amepre.2013.06.017] [Medline: 24139781]
32. Finkelstein EA, Linnan LA, Tate DF, Birken BE. A pilot study testing the effect of different levels of financial incentives on weight loss among overweight employees. J Occup Environ Med 2007 Sep;49(9):981-989. [doi: 10.1097/JOM.0b013e31813c6dcb] [Medline: 17848854]
33. Latham GP, Locke EA. Self-regulation through goal setting. Organ Behav Hum Decis Process 1991 Dec;50(2):212-247. [doi: 10.1016/0749-5978(91)90021-k]
34. Latham GP, Locke EA. Enhancing the benefits and overcoming the pitfalls of goal setting. Organ Dyn 2006;35(4):332-340. [doi: 10.1016/j.orgdyn.2006.08.008]
35. Harrison J, O'Donoghue A, Hill J. The use of pedometers in the workplace to increase physical activity. Br J Card Nurs 2021 Aug 02;16(8):1-4 [FREE Full text] [doi: 10.12968/bjca.2021.0066] [Medline: 38125904]

36. Freak-Poli R, Cumpston M, Albarqouni L, Clemes SA, Peeters A. Workplace pedometer interventions for increasing physical activity. Cochrane Database Syst Rev 2020 Jul 21;7(7):CD009209. [doi: 10.1002/14651858.CD009209.pub3] [Medline: 32700325]
37. Forberger S, Wichmann F, Comito CN. Nudges used to promote physical activity and to reduce sedentary behaviour in the workplace: results of a scoping review. Prev Med 2022 Feb;155:106922. [doi: 10.1016/j.ypmed.2021.106922] [Medline: 34933021]
38. Mackenzie K, Such E, Norman P, Goyder E. Sitting less at work: a qualitative study of barriers and enablers in organisations of different size and sector. BMC Public Health 2019 Jul 04;19(1):884 [FREE Full text] [doi: 10.1186/s12889-019-7148-8] [Medline: 31272484]
39. Mackenzie K, Such E, Norman P, Goyder E. The development, implementation and evaluation of interventions to reduce workplace sitting: a qualitative systematic review and evidence-based operational framework. BMC Public Health 2018 Jul 04;18(1):833 [FREE Full text] [doi: 10.1186/s12889-018-5768-z] [Medline: 29973188]
40. Albulescu P, Macsinga I, Rusu A, Sulea C, Bodnaru A, Tulbure BT. "Give me a break!" A systematic review and meta-analysis on the efficacy of micro-breaks for increasing well-being and performance. PLoS One 2022 Aug 31;17(8):e0272460 [FREE Full text] [doi: 10.1371/journal.pone.0272460] [Medline: 36044424]
41. Danquah IH, Tolstrup JS. Standing meetings are feasible and effective in reducing sitting time among office workers-walking meetings are not: mixed-methods results on the feasibility and effectiveness of active meetings based on data from the "take a stand!" study. Int J Environ Res Public Health 2020 Mar 05;17(5):1713 [FREE Full text] [doi: 10.3390/ijerph17051713] [Medline: 32151044]
42. Biddle SJ, O'Connell SE, Davies MJ, Dunstan D, Edwardson CL, Esliger DW, et al. Reducing sitting at work: process evaluation of the SMArT Work (Stand More At Work) intervention. Trials 2020 May 13;21(1):403 [FREE Full text] [doi: 10.1186/s13063-020-04300-7] [Medline: 32404181]
43. Jakobsen MD, Sundstrup E, Brandt M, Andersen LL. Psychosocial benefits of workplace physical exercise: cluster randomized controlled trial. BMC Public Health 2017 Oct 10;17(1):798 [FREE Full text] [doi: 10.1186/s12889-017-4728-3] [Medline: 29017479]
44. Rush B, Ogborne A. Program logic models: expanding their role and structure for program planning and evaluation. Can J Program Eval 1991;6(2):95-106. [doi: 10.3138/cjpe.6.005]

Edited by A Mavragani; submitted 23.02.24; peer-reviewed by E Breckin; comments to author 09.09.24; revised version received 29.09.24; accepted 10.10.24; published 02.01.25.

Please cite as:
Akksilp K, Rouyard T, Isaranuwatchai W, Nakamura R, Müller-Riemenschneider F, Teerawattananon Y, Chen C
The Physical Activity at Work (PAW) Program in Thai Office Workers: Mixed Methods Process Evaluation Study
JMIR Form Res 2025;9:e57604
URL: https://formative.jmir.org/2025/1/e57604
doi:10.2196/57604
PMID:

Original Paper

ChatGPT's Attitude, Knowledge, and Clinical Application in Geriatrics Practice and Education: Exploratory Observational Study

Huai Yong Cheng[1], MPH, MS, MD

Minneapolis VA Health Care System, Minneapolis, MN, United States

Corresponding Author:
Huai Yong Cheng, MPH, MS, MD
Minneapolis VA Health Care System
1 Veterans Dr.
Minneapolis, MN, 55417
United States
Phone: 1 6124672051
Fax: 1 6127252084
Email: wchengwcheng@gmail.com

Abstract

Background: The increasing use of ChatGPT in clinical practice and medical education necessitates the evaluation of its reliability, particularly in geriatrics.

Objective: This study aimed to evaluate ChatGPT's trustworthiness in geriatrics through 3 distinct approaches: evaluating ChatGPT's geriatrics attitude, knowledge, and clinical application with 2 vignettes of geriatric syndromes (polypharmacy and falls).

Methods: We used the validated University of California, Los Angeles, geriatrics attitude and knowledge instruments to evaluate ChatGPT's geriatrics attitude and knowledge and compare its performance with that of medical students, residents, and geriatrics fellows from reported results in the literature. We also evaluated ChatGPT's application to 2 vignettes of geriatric syndromes (polypharmacy and falls).

Results: The mean total score on geriatrics attitude of ChatGPT was significantly lower than that of trainees (medical students, internal medicine residents, and geriatric medicine fellows; 2.7 vs 3.7 on a scale from 1-5; 1=*strongly disagree*; 5=*strongly agree*). The mean subscore on positive geriatrics attitude of ChatGPT was higher than that of the trainees (medical students, internal medicine residents, and neurologists; 4.1 vs 3.7 on a scale from 1 to 5 where a higher score means a more positive attitude toward older adults). The mean subscore on negative geriatrics attitude of ChatGPT was lower than that of the trainees and neurologists (1.8 vs 2.8 on a scale from 1 to 5 where a lower subscore means a less negative attitude toward aging). On the University of California, Los Angeles geriatrics knowledge test, ChatGPT outperformed all medical students, internal medicine residents, and geriatric medicine fellows from validated studies (14.7 vs 11.3 with a score range of –18 to +18 where +18 means that all questions were answered correctly). Regarding the polypharmacy vignette, ChatGPT not only demonstrated solid knowledge of potentially inappropriate medications but also accurately identified 7 common potentially inappropriate medications and 5 drug-drug and 3 drug-disease interactions. However, ChatGPT missed 5 drug-disease and 1 drug-drug interaction and produced 2 hallucinations. Regarding the fall vignette, ChatGPT answered 3 of 5 pretests correctly and 2 of 5 pretests partially correctly, identified 6 categories of fall risks, followed fall guidelines correctly, listed 6 key physical examinations, and recommended 6 categories of fall prevention methods.

Conclusions: This study suggests that ChatGPT can be a valuable supplemental tool in geriatrics, offering reliable information with less age bias, robust geriatrics knowledge, and comprehensive recommendations for managing 2 common geriatric syndromes (polypharmacy and falls) that are consistent with evidence from guidelines, systematic reviews, and other types of studies. ChatGPT's potential as an educational and clinical resource could significantly benefit trainees, health care providers, and laypeople. Further research using GPT-4o, larger geriatrics question sets, and more geriatric syndromes is needed to expand and confirm these findings before adopting ChatGPT widely for geriatrics education and practice.

(JMIR Form Res 2025;9:e63494) doi:10.2196/63494

KEYWORDS

ChatGPT; geriatrics attitude; ageism; geriatrics competence; geriatric syndromes; polypharmacy; falls; aging, older adults

Introduction

Background

ChatGPT stands for Chat Generative Pretrained Transformer and was developed by an artificial intelligence (AI) research company, OpenAI. It is an AI chatbot technology that can process our natural human language and generate a response. ChatGPT was released to the public by OpenAI on November 30, 2022, and surpassed 1 million users in just 5 days. This performance set a record for the second fastest-growing user base after that of Threads. ChatGPT currently has >180 million users (as of October 26, 2024) [1]. ChatGPT users have increased exponentially, and it has been increasingly used in medicine, resulting in 4625 publications on PubMed (as of October 26, 2024) [2]. Notably, ChatGPT was not programmed originally to be used for medical education and clinical application [1] despite reports of its wide application for medical education [3-7] and clinical practice [5-12]. ChatGPT has been studied in multiple specialties and subspecialties, such as psychiatry [13], radiology [14], and cardiology [15]. Its rapid progress in such a brief time has raised challenges, concerns, and limitations, including privacy, ethnic and racial bias, and legal risks [16-21].

Any application of ChatGPT to medical education and clinical practice could be promising but needs to be well designed and investigated. In the context of a growing aging population, geriatrics education, geriatrics workforces, and geriatrics practice have been facing significant challenges for the last 4 decades [22-26]. ChatGPT could offer an opportunity to improve geriatrics education and clinical practice [27-30]. Preliminary studies have shown promise [27-30]. For example, older adults often have polypharmacy issues [31], and providers review and deprescribe medications [32]. A specifically trained large language model may provide useful clinical support in polypharmacy management for primary care physicians [29]. ChatGPT has scored better on general and theoretical questions in geriatrics than on complex decisions and end-of-life situations, with the lowest scores related to diagnosis and performing complex tests [30]. Overall, the application of ChatGPT to geriatrics education and clinical practice is sporadic and needs to be expanded and further investigated. Therefore, this study was designed to determine whether we can trust ChatGPT to be used in geriatrics education and clinical practice using 3 distinct approaches.

The first approach was to evaluate any age-biased outputs generated by ChatGPT. The aging narrative spans >210 years [33]. The term ageism, coined by Robert Butler in 1969, refers to age-based stereotyping, prejudice, and discrimination [34]. Self-perception of aging is another emergent concept related to ageism [35]. Ageism is global and prevalent in daily life [34,36,37], health care systems [34,38], and the news media [39,40]. It is present in various medicine specialties such as oncology [41] and cardiology [42] and was persistent during the COVID-19 pandemic [43,44]. Ageism is significantly associated with poor health outcomes and quality of life [34,35]. Interventions to reduce ageism have been widely studied [45,46]. ChatGPT has been found to exhibit privacy, ethnicity, gender, and racial bias and entail legal risks [16-21]. For example, ChatGPT recommends fewer female than male ophthalmologists [47]. However, whether ChatGPT generates age-biased outputs has not been reported. This study aimed to demonstrate whether ChatGPT exhibits ageism by testing its geriatrics attitude using a validated University of California, Los Angeles (UCLA) geriatrics attitude instrument [48,49] and comparing its results with those obtained by medical students residents, and geriatrics medicine fellows from published articles

The second approach was to evaluate the geriatrics knowledge of ChatGPT. ChatGPT has passed the United States Medical Licensing Examination (USMLE) [50,51] and many other medicine exams [52-56], demonstrating its competence in medical knowledge. However, it has been less studied in geriatrics. For example, in one Geriatrics and AI in Spain project in Spain [30], ChatGPT was prompted with 10 questions about geriatric medicine. Compared to 130 physicians who answered the questionnaire, ChatGPT scored better on general and theoretical questions than on complex decisions and end-of-life situations, with the lowest scores for diagnosis and performing complex tests [30]. AI is likely to be incorporated into some areas of geriatric medicine, but it still presents significant limitations, mainly in complex medical decision-making [30]. This study was designed to demonstrate the geriatrics competency of ChatGPT by examining its performance on the validated UCLA geriatrics knowledge test [48,49,57] and comparing it with that of medical students, internal medicine residents, and geriatric medicine fellows from previously published articles in the literature.

The third approach was to evaluate ChatGPT's knowledge of 2 geriatric syndromes and its clinical application to them (polypharmacy and falls). Multiple previous studies have shown that ChatGPT performs well on clinical vignette questions [58-64]. For example, ChatGPT achieved 71.7% accuracy overall on 36 clinical vignettes from the Merck Sharpe and Dohme clinical manuals and impressive accuracy in clinical decision-making [58]. In another study, GPT-4o was queried for diagnoses and management plans with 20 physician-written clinical vignettes in otolaryngology, showing high agreement with physicians [59]. In addition, ChatGPT demonstrated appreciable knowledge and interpretation skills in psychiatry through 100 clinical case vignettes [60]. In another study, 33 physicians across 17 specialties generated 284 medical questions. ChatGPT generated largely accurate information in response to diverse medical queries as judged by academic physician specialists, with improvement over time [61]. ChatGPT also offered accurate recommendations on managing hypertension based on clinical practice guidelines [62]. ChatGPT provided accurate recommendations for 8 out of 10 clinical scenarios in cardiology [63]. In another study, GPT-4o generated a good rehabilitation plan for a stroke case from a textbook [64]. Finally, several studies have reported the application of

ChatGPT to predict common drug-drug interactions [65], managing polypharmacy [29], clinical pharmacy [66-68], and medical pharmacology as a self-learning tool [69]. Taken together, all previous studies suggest that ChatGPT could potentially be applied to complex geriatric syndrome vignettes with comprehensive clinical questions. This study was designed to demonstrate whether ChatGPT has solid geriatrics knowledge and can apply it to 2 common complex geriatric syndrome vignettes (polypharmacy and falls) by responding to comprehensive questions.

Study Objectives

We designed these 3 distinct approaches to provide evidence of whether ChatGPT can be trusted to be potentially applied to geriatrics education and clinical practice as an assistive tool.

Methods

This observational study used the validated UCLA geriatrics attitude instrument [48,49] and geriatrics knowledge test [57].

The Geriatrics Attitude Instrument

The geriatrics attitude instrument comprises 16 questions (Multimedia Appendix 1) [48,49]. In total, 6 of these statements reflect positive attitudes toward aging, such as "Most old people are pleasant to be with" (question 1). A total of 10 statements reflect negative attitudes toward aging, such as "Treatment of chronically ill old patients is hopeless" (question 11). Responses are graded on a Likert scale from 1 (*strongly disagree*) to 5 (*strongly agree*). The mean total score of ChatGPT was calculated and compared to scores from medical students, internal medicine residents, and geriatric medicine fellows based on previously published studies in the literature. Subtotals for positive and negative attitudes toward aging were also calculated to compare them to those in previously published studies. For statements reflecting a positive attitude toward aging, higher scores indicated a more positive attitude toward aging. For statements reflecting a negative attitude toward aging, lower scores indicated a less biased attitude toward aging.

The Geriatrics Knowledge Test

The validated UCLA geriatrics knowledge test contains 18 questions (Multimedia Appendix 2) [57]. In total, 8 questions are true-or-false statements, such as "Most older people are living in nursing homes" (question 4; false). A total of 10 questions are short clinical vignettes, such as "A 78-year-old nursing home resident with mild dementia associated with Alzheimer's disease is best determined by their ability to understand treatment options" (question 9; correct answer: B).

To minimize unfounded guessing, the following scoring system was used: +1 for a correct answer, –1 for a wrong answer, and 0 for "don't know." Therefore, the total score ranged from −18 to +18 [57]. Correct answers were determined by the author based on the literature.

ChatGPT Inputs and Outputs

GPT-3.5 was prompted to undertake the validated UCLA geriatrics attitude instrument and geriatrics knowledge test. Prompts were derived from original published papers [48,49,57]. Each prompt was repeated 4 times to evaluate the consistency and obtain an average response. In addition, ChatGPT was prompted to respond to 2 common geriatric syndrome vignettes (polypharmacy and falls). These 2 vignettes with questions were used in the author's previously published geriatrics curricula [70,71] and are described fully in the following sections.

The Polypharmacy Vignette

This vignette was used from a previously published workshop of prescribing and deprescribing [70]:

> *Ms. Smith, an 85-year-old white woman, has HTN, CAD, HFpEF, A-Fib, DM (II), HLD, hypothyroidism, gout, OA. On Warfarin 2.5 mg/d, furosemide 20 mg/d, Glyburide 5 mg/d, Spironolactone 50 mg/d, Amiodarone 300mg/d, Carvedilol 6.25 mg twice a day, Lisinopril 40 mg/d, Lipitor 80 mg/d, Levothyroxine 250 mg/d, Allopurinol 100 mg three times a day, acetaminophen 500 mg two tabs q6h for joint pain, and several supplements. Her pulse is 62 beat/min. Her BP is 120/70 (sit) and 98/64 (upright), weight 60 kg. No JVD, HR regularly, no murmur, few crackles on lung bases. Abdomen exam is benign, + pitting edema of ankles. Serum Cr is 1.2 mg/dL and Bun is 15 mg/dL. INR is 3.5. LDL is 50. albumin 4, Hb1ac is 6.1. TSH is 2. EF is 60%.*

ChatGPT was prompted with the polypharmacy vignette following 2 steps. First, ChatGPT was prompted to answer 4 general questions on the appropriate use of medications to examine its geriatric pharmacology knowledge: (1) "Which of the commonly used 8 drugs are potentially inappropriate medication (PIM) and should be avoided in older adults?" (2) "What is PIM in older adults?" (3) "How to identify PIM in older adults?" (4) "What to do with PIM?" Second, ChatGPT was prompted to answer 4 additional specific questions related to the polypharmacy vignette to evaluate its application of geriatric pharmacology to a clinical vignette: (1) "Any drug-drug interactions for this patient?" (2) "Any drug is ineffective for this patient?" (3) "This patient was taking so many medications. Any way to reduce the number of her medications?" (4) "Any drug-disease interaction?"

The Fall Vignette

This vignette was used from a previously published workshop of prescribing and deprescribing [71]:

> *An 82-year-old community-dwelling woman with PMH of multiple falls, right hip fracture, Dx of hypertension, coronary heart disease, knee arthritis, came to my clinic after she fell after her meal at home this morning. She felt pain of R head. She had a cane and walker at home but didn't use that often. She thought she didn't need it. She has taken care of her husband for several years. Her medication included Doxazosin (Cardura) and diltiazem (Cardizem), and acetaminophen as needed. I asked many questions related to her fall. Her vital signs were normal without orthostatic hypotension. Her cardiovascular and neurological examinations were benign. Her mini-mental status examination (MMSE) was 30/30.*

> *She had independent activities of daily living (ADL). Her geriatric depression scale (GDS) was 0/15. She passed timed up and go test. Because of her right black eye and pain of right head, I sent her to UVA to rule out orbital fracture. Please ask the following questions: List all potential fall risk factors for this patient What are three fall screening questions you need to ask for elderly patients in an outpatient setting based on 2010 fall prevention guideline you want to follow for this patient? Which one of three pathways in figure 1 from 2010 fall prevention guideline you want to follow for this patient? List all physical examinations that you think are important to perform for this patient. List your recommendations as many as you think that could help her to prevent her falling.*

ChatGPT was prompted with the fall vignette following 3 steps. First, ChatGPT was prompted to answer 1 general question on fall risks in older adults. Second, ChatGPT was prompted to answer five pretest questions on fall prevention in older adults: (1) "When an older person falls, should you ask the following questions (one or more or none)? the circumstances of fall, mental status, medication changes, vision change, drinking"; (2) "Which of the following illnesses is a fall risk (one or more or none)? Pneumonia, Parkinson's disease, Diabetes mellitus, Dementia, Orthostatic hypotension"; (3) "Which of the following medications are not associated with an increased fall risk (one or more)? Levothyroxine, Diazepam, Oxycodone, Amitriptyline"; (4) "Which ones of the following are true (one or more or none)? An elderly person may become so fearful of falling that they restrict his or her mobility Most falls are associated with significant injuries Falls is a leading cause of accidental death in older population About 5% community-dwelling elderly person falls each year Survivors of fall-related hip fracture are rarely institutionalized Exercise improves function and reduces fall risk and injurious falls"; and (5) "Which of the following is NOT an environmental fall risk? (only pick one answer) Throw rug, freshly waxed kitchen floor, Grab bars, Electrical cord lying on the floor." Finally, ChatGPT was prompted to answer 5 additional questions to examine the application of fall prevention knowledge to the fall vignette: (1) "List all potential fall risk factors for this patient"; (2) "What are three fall screening questions you need to ask for elderly patients in an outpatient setting based on 2010 fall prevention guideline you want to follow for this patient?"; (3) "Which one of three pathways in figure 1 from 2010 fall prevention guideline you want to follow for this patient?"; (4) "List all physical examinations that you think they are important to perform for this patient"; and (5) "List your recommendations as many as you think that could help her to prevent her falling."

Evaluation of ChatGPT Outputs

ChatGPT outputs were collected and analyzed by the author. The author judged the correctness of the ChatGPT outputs on the polypharmacy and fall vignette questions based on evidence from clinical practice guidelines, systematic reviews, and other types of evidence in the literature. The descriptive analysis was performed using SPSS (version 28; IBM Corp).

In summary, we designed these 3 distinct approaches to provide evidence on whether we can trust ChatGPT to be potentially applied to geriatrics education and clinical practice as an assisting tool.

Ethical Considerations

This study was not human participant related, including the secondary data analysis. No informed consent was needed.

Results

Geriatrics Attitude

The total geriatrics attitude score of ChatGPT was significantly lower than that of trainees (medical students, internal medicine residents, and geriatric medicine fellows) from validation [49,57] and follow-up studies [48] (Multimedia Appendix 3). However, the mean subscore on positive geriatrics attitude of ChatGPT was higher than that of the trainees (medical students, internal medicine residents, and neurologists; Multimedia Appendix 4), where a higher score is better. The higher subscore on positive geriatrics attitude indicates a better attitude toward aging. Conversely, the mean subscore on negative geriatrics attitude of ChatGPT was lower than that of the trainees and neurologists, where lower subscores are better (Multimedia Appendix 4). The lower subscore on negative geriatrics attitude indicates a less age-biased attitude toward aging. Individual responses to 14 geriatrics attitude statements by ChatGPT, trainees, and neurologists are shown in Multimedia Appendix 4. The subscores of the comparison group in Multimedia Appendix 4 were based on previously published studies [72-74]. Notably, responses to statements 15 and 16 were rarely reported in previously published studies (Multimedia Appendix 4).

Of the 14 geriatrics attitude statements (Multimedia Appendix 4), 5 (36%) were positive, and 9 (64%) were negative. Regarding the positive geriatrics attitude statements, ChatGPT had 1 lower positive geriatric attitude response (indicating a less positive geriatrics attitude) and 5 similar or better positive geriatric attitude responses compared to medical students and residents (indicating a similar or better positive geriatrics attitude). Regarding the negative geriatrics attitude statements, ChatGPT had 1 similar negative geriatric attitude response and 8 better negative geriatrics attitude responses than medical students and residents, indicating a less negative geriatrics attitude.

The response from ChatGPT to the first prompt of the geriatrics attitude questions did not follow the Likert-scale format (1-5; 1=*strongly disagree*; 5=*strongly agree*; Multimedia Appendices 3 and 4) as ChatGPT provided only comments. The response to the third prompt included both the Likert-scale format (1-5; 1=*strongly disagree*; 5=*strongly agree*) and comments (Multimedia Appendix 4).

Geriatrics Knowledge

The total score of ChatGPT (mean 14.25) was significantly higher than the scores from all trainees (medical students, residents, and internal medicine fellows) in the validation studies [57] (mean 11.3; Multimedia Appendix 3) and was also significantly higher than that of first-year (mean 9.9) and second-year (mean 9.5) medical students (Multimedia Appendix

3) in the follow-up studies [48]. It was slightly higher than the scores of third-year medical students (mean 13.6) and internal medicine interns (mean 13.2), slightly lower than the scores of second- and third-year internal medicine residents (mean 14.7 and 14.9, respectively), and significantly lower than the scores of geriatric medicine fellows (mean 17.5) in the follow-up studies [48]. ChatGPT's performance on geriatrics knowledge in the first prompt was lower than that on the following 3 repeat prompts. For the first prompt, ChatGPT provided rationales without selecting options (Multimedia Appendix 4). For the third prompt, ChatGPT selected options and provided rationales for its choices (Multimedia Appendix 4). For the second and fourth prompts, ChatGPT select options without providing rationales (Multimedia Appendix 4).

To make it easy for the reader to replicate the results of this study, Multimedia Appendix 5 shows a few examples of prompts on geriatrics attitude and knowledge test questions and outputs by ChatGPT (all original outputs of ChatGPT are available upon request).

The Polypharmacy Vignette

Overall, ChatGPT performed well on the polypharmacy vignette involving a woman aged 85 years (Multimedia Appendix 6). ChatGPT provided appropriate responses to 4 general drug therapy questions and moderate responses to 4 specific drug therapy questions based on the vignette. ChatGPT correctly identified 5 drug-drug interactions and suggested deprescribing. However, it missed an ineffective medication (a supplement), 1 drug-drug interaction, and 3 drug-disease interactions (Multimedia Appendix 6). Despite the patient not having lung disease, ChatGPT provided 2 irrelevant drug-disease interactions specific to the vignette (ChatGPT hallucination; Multimedia Appendix 6).

The Fall Vignette

ChatGPT performed well on the fall vignette involving a woman aged 82 years (Multimedia Appendix 7). ChatGPT correctly summarized 10 common fall risks in older adults and correctly answered 3 out of 5 pretest questions. In the remaining 2 pretest questions, ChatGPT missed a few correct responses. In the fall vignette, ChatGPT provided appropriate responses to all 4 prompts (Multimedia Appendix 7). ChatGPT recognized most fall risk factors and responded perfectly to fall screening questions following the latest Centers for Disease Control and Prevention (CDC) fall guidelines even though the prompt specified the 2010 fall prevention guidelines. ChatGPT accurately followed the 3 pathways from the 2010 fall guidelines. It missed a few physical examinations. ChatGPT provided almost all recommended fall prevention strategies, with only a few omissions.

Discussion

Principal Findings

With the increasing use of ChatGPT in medical education [3-7] and clinical practice [5-12], the aim of this study was to evaluate whether we can trust ChatGPT to be used in geriatrics. The major findings of this study demonstrated that ChatGPT had less age-biased output than the trainees, outperformed the trainees on the validated UCLA geriatrics knowledge test, and reasonably applied geriatrics knowledge to 2 common geriatric syndrome vignettes (polypharmacy and falls). The preliminary findings of this study are promising and demonstrate that ChatGPT could be trusted and used as an assistant tool in geriatrics practice and education via 3 approaches, which are fully discussed in the following paragraphs.

The first approach was to evaluate ChatGPT's performance on a validated UCLA geriatrics attitude instrument [48,49]. To the best of our knowledge, this is the first study to demonstrate that ChatGPT is less likely to generate age-biased outputs, which contrasts with a few ethical and other concerns regarding ChatGPT [16-21]. This finding is significant because ageism is prevalent and associated with multiple poor outcomes, including health outcomes [34,35]. Various instruments exist to measure geriatrics attitude to assess ageism [34,35], with the UCLA geriatrics attitude instrument [48,49] being one of these validated tools [75], which we used in this study [48,49]. The UCLA geriatrics attitude instrument includes 16 items [48,49], 6 of which refer to a positive geriatrics attitude (eg, "most old people are pleasant to be with") where a higher score is better, and 10 of which refer to a negative geriatrics attitude where a lower score is better. Previous studies using the UCLA geriatrics attitude instrument have primarily used the total score on the geriatrics attitude scales without calculating and reporting the mean subscores on positive and negative geriatrics attitudes [48,49], which makes it difficult to interpret the significance of geriatrics attitude. However, a few studies have reported individual responses to the UCLA geriatrics attitude instrument, including positive and negative geriatrics attitude [48,49], for which we were able to calculate the subscores for comparison. We recommend that the subscores of the UCLA geriatrics attitude instrument should be reported to assess positive and negative geriatrics attitude. Overall, ChatGPT's individual responses to the UCLA geriatrics attitude statements were better than those of medical students and residents, with higher scores for positive attitude and lower scores for negative attitude (Multimedia Appendix 4), suggesting that ChatGPT has less age-biased outputs and could be trusted from an ethical perspective. This study suggests that ChatGPT is better than the news media and humans in providing less age-biased information [39,40]. However, a reliable and valid instrument with which to quantify modern medical student attitudes toward older people has not yet been developed. An adaptation of the Aging Semantic Differential scale for contemporary use has been recommended [76]. This study can be repeated using the Aging Semantic Differential scale to see whether ChatGPT still generates less age-biased outputs in the future.

The second approach was to evaluate ChatGPT's geriatrics knowledge. This study is the first to assess ChatGPT's performance on a validated UCLA geriatrics knowledge test [49,57]. ChatGPT has a substantial knowledge base and has been reported to pass the USMLE [50,51] and perform well on many other examinations [52-56]. Previous studies have suggested that ChatGPT could significantly impact medical education [3-7] and clinical practice, including clinical decision-making [5-12]. For example, a recent systematic review and meta-analysis on the performance of ChatGPT in medical

examinations showed that ChatGPT has been evaluated in multiple specialties, including plastic surgery, ophthalmology, neurosurgery, orthopedics, diabetes, gastroenterology, radiology, cardiology, dermatology, and anesthesia [61]. It can be reasonably believed that ChatGPT has good geriatrics knowledge. This study supported this belief and demonstrated that ChatGPT performed better than the average trainee in the validation studies (mean score 14.7 vs 11.3) and follow-up studies (mean score 14 vs 13.3; Multimedia Appendix 3). ChatGPT's performance was better than that of medical students and first-year internal medicine residents (PGY 1 in Multimedia Appendix 3), comparable to that of second- and third-year residents (PGY 2 and PGY 3 in Multimedia Appendix 3), but lower than that of geriatric medicine fellows. This suggests that ChatGPT could be a valuable supplemental resource for health professional trainees, health care providers, and laypeople. Given that GPT-4o has more parameters, it is reasonable to believe that GPT-4o could perform better on the geriatrics knowledge test. The UCLA geriatrics knowledge test only has 18 questions [57] and much less questions than the USMLE tested in previous studies [50,51]. Future studies should evaluate GPT-4o's performance on more multiple-choice questions from geriatrics board examinations to confirm the findings of this study before adopting ChatGPT for geriatrics education and clinical practice.

The third approach was to evaluate ChatGPT's responses to 2 common geriatric syndrome vignettes—polypharmacy and falls in older adults—from the author's previously published geriatrics curricula [70,71]. This study demonstrated that ChatGPT performed well on these 2 geriatric syndrome vignettes (Multimedia Appendices 6 and 7), which was promising and will be fully discussed in the following paragraphs.

For the polypharmacy vignette, ChatGPT was able to provide systematic recommendations for polypharmacy in older adults, which has not been investigated before. In this study, ChatGPT provided appropriate general principles for prescribing and deprescribing; identified PIMs; and suggested resources and criteria such as the Beers criteria [77], the Screening Tool of Older Persons' Prescriptions [75], the Drug Burden Index [78], and the Medication Appropriateness Index [79] (Multimedia Appendix 6). In addition, ChatGPT was able to provide 5 appropriate ways to manage PIMs in older adults (Multimedia Appendix 6) that are consistent with the evidence from geriatrics practice guidelines and prescribing principles [75,77-80]. In particular, this study demonstrated that ChatGPT can identify all 8 exemplary PIMs (Multimedia Appendix 6). These medications should be avoided in older adults [75,77-80]. ChatGPT was able to identify 4 drug-disease interactions but missed 3 drug-disease interactions in the polypharmacy vignette (Multimedia Appendix 6). This suggests room for improvement in training ChatGPT further. ChatGPT also identified drug-drug interactions accurately in the polypharmacy vignette (Multimedia Appendix 6), which was different from a previous study in which ChatGPT identified 39 out of 40 drug-drug interactions when it was given 40 pairs of drugs [65]. Therefore, the question is what the implications are of what ChatGPT can do for polypharmacy. Previous studies have shown that within the last 10 years medical students, postgraduate residents, primary care providers, and pharmacists were unaware of PIMs and the standard guidelines for older adults, such as the Beers criteria [81-84]. This study has shown that ChatGPT is more knowledgeable than humans. Both geriatricians and general internists still prescribed 7.2% to 8.7% of PIMs, respectively [84].

The persistent prevalence of PIMs and polypharmacy in older adults was well-reported [31,32]. What ChatGPT demonstrated its good knowledge and ability to do with polypharmacy in this study suggests that ChatGPT has great potential to be used by trainees and providers as an assistant tool. One suggested example will be to include an older patient's medical history and a list of all their medications in the ChatGPT prompt used in this study for medication review and to generate ChatGPT outputs which can be reviewed by providers at Clinic or other clinical settings. This will help nurses, other providers, and trainees at the point of care. It could help laypeople self-check PIMs. It could also help health professional trainees in the self-study of geriatric pharmacology.

Falling is another significant US and global problem [85,86] and one of the common geriatric syndromes in older adults [87]. ChatGPT has not been tested to provide systematic fall prevention recommendations for an older adult. This study demonstrated for the first time that ChatGPT has good knowledge of fall prevention as it performed well on pretests and provided a comprehensive summary of 10 common fall risks in older adults and specific recommendations in the fall vignette (Multimedia Appendix 7).

The 2024 US Preventive Services Task Force fall prevention guidelines listed several fall risks, including age; history of falls; cognitive and sensory deficits; presence of acute and chronic medical conditions; certain medications; environmental or occupational hazards; home or neighborhood features and alcohol or drug use; and impairments in mobility, gait, and balance [87]. The 2023 CDC Stopping Elderly Accidents, Deaths, and Injuries also has a long list of fall risks [88]. In addition, a systematic review and meta-analysis showed evidence-based fall risk factors among the aging population [89]. This study showed that ChatGPT identified most of fall risks for the fall vignette (Multimedia Appendix 7) consistent with the aforementioned fall prevention guidelines [87,88] and systematic review [89] except for missing risky behaviors such as standing on a chair instead of a step stool and loss of sensation in the feet.

This study also used 5 pretests to further evaluate ChatGPT's knowledge of fall prevention (Multimedia Appendix 7). ChatGPT generated responses consistent with the guidelines [88,89] and a systematic review [90]. In pretest 1, ChatGPT provided 5 good recommendations with reasoning for taking the medical history of an older adult. In pretest 2, ChatGPT correctly identified several medical conditions as fall risks, including Parkinson disease, dementia, and orthostatic hypotension but missing pneumonia and diabetes mellitus. This indicates that ChatGPT has some limitations. In pretest 3, ChatGPT correctly identified diazepam, oxycodone, and amitriptyline as fall risks [75,77-79] and levothyroxine as not a fall risk. In pretest 4, ChatGPT underestimated the prevalence

of falls in community-dwelling older adults (5%) but correctly answered most other true and false questions about fall risks and consequences. In pretest 5, ChatGPT correctly answered all questions about environmental fall risks and identified grab bars as not a risk. Overall, ChatGPT performed very well on 5 pretests with a few mistakes, indicating good knowledge of fall prevention for older adults.

To further test ChatGPT's clinical application to individual patients, the fall vignette was used to prompt ChatGPT. ChatGPT correctly identified 6 common fall risk factors except for potential home environment factors, consistent with the evidence from fall prevention guidelines [88,89] and a systematic review [90]. ChatGPT impressively recognized older caregiver burden, hypertension, and coronary heart diseases as fall risks, also consistent with the evidence from fall prevention guidelines [88,89] and a systematic review [90]. In response to fall screening questions, ChatGPT was correct if the 2023 CDC guidelines were used [89] but incorrect if the 2010 American Geriatrics Society guidelines were used [91]. This study indicates that ChatGPT can use the latest guidelines, although specific prompts might be unclear to ChatGPT. Regarding physical examinations relevant to the fall vignette, ChatGPT missed several but recommended multiple appropriate physical examinations [89,91], indicating a reasonably good performance. In the final prompt on fall prevention recommendations for the fall vignette, ChatGPT provided 6 appropriate fall prevention recommendations, including home environment–related prevention despite previously missing the home environment as a fall risk [88-91]. It was notable that the home environment was not on the list of fall risks in ChatGPT's response to the first prompt. However, ChatGPT correctly answered pretest 5 (home environment fall risk) and recommended home environment–related prevention for the patient in the fall vignette (Multimedia Appendix 7). This suggests some discordance between the fall assessment and action plans by ChatGPT, indicating a possibly different thinking and decision process within ChatGPT.

The crucial question is whether ChatGPT can be helpful and necessary in screening for fall risks and providing fall prevention recommendations to older adults in daily practice. Providers need knowledge and willingness to screen for fall risks in older adults, but multiple studies have shown that providers often lack fall prevention knowledge and are unwilling to screen for fall risks [92-96]. For example, one study showed that emergency providers lacked knowledge on which patients to be screened and were unwilling to spend more than a few minutes on screening for fall intervention [95]. Another study revealed that only 14% of providers at accountable care organizations were aware of the CDC's fall risk assessment algorithm (Stopping Elderly Accidents, Deaths, and Injuries) [93]. Furthermore, 43% of primary care providers did not agree that they had the expertise or time to perform fall risk assessments, and only a small percentage billed for fall risk screening despite being aware of Medicare reimbursement [96]. In addition, general practitioners were often unaware of their frail patients' fall history or fear of falling, with most patients not receiving fall prevention care. Less than 40% of providers asked most or all their older patients if they had fallen in the previous 12 months, and less than a quarter referred their patients to physical therapists for balance or gait training [92]. A recent CDC report showed some improvement, but gaps remain. Only 20% of providers were aware of any injury prevention resources, and a higher percentage screened for fall risk when older adults presented with specific concerns [94].

Given the high prevalence of fall among older adults and poor knowledge of fall prevention among providers, this study demonstrated that ChatGPT has better knowledge and ability to apply the fall prevention guidelines than many providers, suggesting that it can assist providers in assessing fall risks and preventing falls in older adults. For instance, entering an older patient's history into ChatGPT and reviewing the outputs generated can help providers identify patients with fall risk at the point of care. ChatGPT can also be used as an assistant tool to help health professional trainees with self-study.

Limitations and Quality Assessment of This Study

There are serval limitations to this study, which will be discussed in this section. ChatGPT's responses are based on the data that it was trained on, which may not include the most current geriatrics research, best practices, or region-specific guidelines, potentially limiting the trustworthiness of its advice. This study did not account for the quality of the information sources that ChatGPT draws from as ChatGPT cannot differentiate between authoritative and nonauthoritative content in real time like all other ChatGPT-based research. Despite GPT-4o having more data for training without releasing its data sources, this will not solve this basic problem. This study indicates that applying ChatGPT to geriatrics practice and education should be cautious and it should be used as an assistant tool only. ChatGPT will never or is unlikely to replace clinicians. As this study evaluated ChatGPT based on geriatric syndrome vignettes, it does not fully account for the complexities and nuances of real-world clinical decision-making and patient interactions, where additional context, history, and human judgment play critical roles. In addition, ChatGPT's responses might be overly generalized as it lacks the ability to perform individualized assessments or make nuanced clinical decisions based on patient-specific data. This is similar to vignette-based simulation education. This indicates that ChatGPT will not make a decision for someone but could assist with the decision-making. The effectiveness of ChatGPT's responses varies significantly depending on how well the user phrases the questions or provides necessary information. The input provided in this study was based on the author's geriatrics practice experience for >20 years, which might not mimic the way in which other clinicians would interact with the tool. Therefore, the results of this study may not be generalizable. Prompt engineering is developing to help improve the quality of the input. However, a standardized prompts among different providers can be very challenging. The author suggests practicing prompts to improve human-ChatGPT interaction and obtain reliable outputs from ChatGPT.

ChatGPT lacks clinical experience and the ability to engage in human clinical judgment, whereas it can process vast amounts of medical information. This may limit its application in critical decision-making scenarios or nuanced treatment planning for

complex geriatrics cases. Providers must use their clinical judgment integrated with clinical circumstances and not fully depend on the outputs of ChatGPT.

ChatGPT might face ethical and legal consequences. ChatGPT or AI hallucinations could cause ethical and legal risks. After the release of GPT-4o, ChatGPT often says, "I am AI and not clinician and cannot make decisions for you." It seems that ChatGPT developers recognize the potential legal and ethical consequences and protect ChatGPT.

Human-AI interaction is complex. Providers should be cautious in interpreting ChatGPT outputs. ChatGPT outputs should be interpreting with evidence-based clinical guidelines, systematic reviews, and randomized controlled trials shown in the Comments from the author columns in Multimedia Appendices 6 and 7. Whether the ChatGPT outputs are appropriate in this study should be judged and verified by a group of geriatricians

In addition, this study has several other limitations. This study used GPT-3.5. GPT-4o might produce better results. Comparisons between ChatGPT and human performance on geriatrics attitude and knowledge were based on previously reported results in the literature many years ago. A new study assessing geriatrics attitude and knowledge among current medical students, residents, and geriatric medicine fellows is needed to confirm the findings of this study. ChatGPT produced some errors, including ChatGPT hallucinations such as unrelated drug-disease interactions in Multimedia Appendix 6, consistent with previous reports [97,98]. We should be aware that ChatGPT can make mistakes, just like humans. ChatGPT should not be used alone in geriatrics practice and education at this point. ChatGPT needs to be pretrained specifically for medical education and clinical practice.

Like any other research, systematic appraisal of the current ChatGPT research is needed. A preliminary checklist, Model, Evaluation, Timing, Range and Randomization, Individual Factors, Counts, and Specificity of Prompts and Language (METRICS), to standardize the design and reporting of studies on generative AI–based models in health care education and practice has been recently developed [99]. The current study was started and completed before its development. However, it is worthwhile to use METRICS to evaluate the quality of this study. The following seven domains of METRICS will be used:

1. Model—ChatGPT was described and used.
2. Evaluation—both subjective and objective evaluation were described and used.
3. Timing—the study date was documented, but the duration was not documented; transparency—the geriatrics attitude instrument and knowledge test, the vignettes, and prompts were well described, and the outputs were exactly presented.
4. Range—the topics were well described, including geriatrics attitude and knowledge; randomization was not used.
5. Individual—there was subjective involvement with ChatGPT, such as judging the correctness of its outputs. This study was conducted by a single investigator.
6. Count—a total of 4 prompts were used. This study did not have a sample size.
7. Specificity of the prompts or language—for geriatrics attitude and knowledge, the prompts were from previously published studies. It is unclear whether they are appropriate for ChatGPT. For the 2 vignettes on polypharmacy and fall, the prompts were from previously published curricula. It is also unclear whether they are appropriate for ChatGPT.

Overall, the quality of this study is reasonably good based on METRICS. METRICS should be used in the reporting of any AI-based research in all journals.

In summary, this study took 3 distinct approaches to demonstrate the trustworthiness of ChatGPT to be used as an assistant tool in geriatrics practice and education. The implications of this preliminary study for geriatrics practice and education are further discussed in the following section.

Implications for Geriatrics Practice and Education

This study showed that ChatGPT outputs are not age-biased using the validated geriatrics attitude test and that ChatGPT performed well on the validated geriatrics knowledge test and on applied geriatrics knowledge tests (2 cases of common geriatric syndromes). These findings suggest that ChatGPT could potentially help geriatrics practice and education as an assistant tool in numerous ways. It is very important to know that ChatGPT is algorithm based. Clinical practice often uses algorithm- and pathway. Therefore, the underlying rationale for ChatGPT and clinical practice is similar. However, ChatGPT is not intended to replace physicians and other providers.

ChatGPT can be used as an assistant geriatrics tool and tutor to support self-learning and immediate feedback and self-assessment. For example, when trainees are studying geriatric syndromes, they can obtain responses on geriatric syndromes from ChatGPT. Therefore, ChatGPT could be incorporated to supplement the existing program in geriatrics education and support continuing medical education. ChatGPT can be potentially to be integrated to clinical reasoning and decision in a timely fashion and potentially improve the point of care in all clinical settings. For example, ChatGPT can help identify individual older patients who might have fall risks and drug-drug interactions.

ChatGPT can be used as an assistant or autonomous clinical tool to alleviate the geriatrics workforce crisis. For example, a nongeriatrician health provider could look for answers to clinical geriatrics questions, such as recommendations for preventing falls and reducing polypharmacy in older patients, from ChatGPT and help answer their patients' questions in a timely fashion. This is particularly helpful for rural practice where geriatrics consults are rare or nonexistent. This will improve physicians' and other providers' efficiency and accuracy in the care of older patients and provider-patient interactions.

It is expected that ChatGPT can help with post-visit summaries by providing patient education materials, such as home safety and environment modification for fall prevention. ChatGPT can also draft Clinic letters such as providing lab results and their interpretation to the patient in the Clinic letter.

ChatGPT can help older patients with self-screening or initial self-assessment when they are sitting in a clinic waiting room. For example, they can input their medical history into ChatGPT

such as age, physical function medications, and chronic conditions to ask whether they are at higher risk of fall.

Importantly, GPT-3.5 is free to use and can be cost-saving for providers and the health care system.

Finally, with more data training, ChatGPT will be more reliable and helpful in geriatrics practice and education.

Comparison With Previous Work

To the best of our knowledge, health professional trainees and providers but not ChatGPT have been assessed for geriatrics attitude [48,49]. ChatGPT has been assessed for geriatrics knowledge [28-30]. For example, ChatGPT was asked to answer 10 geriatrics questions [30]. The correctness of the outputs by ChatGPT were rated by a group of geriatricians [30]. In contrast, in this study, the correctness of the outputs by ChatGPT was determined by the author only, which could lead to errors. In another study, ChatGPT was prompted with 1 simple question [28]. The questions in the aforementioned studies were not validated. However, this study used validated UCLA geriatrics knowledge tests [57,72-74]. Several previous studies have demonstrated that ChatGPT has geriatric pharmacology knowledge [29,65-69]. For example, ChatGPT could identify drug-drug interactions from 40 drug-drug pairs [65]. In another study, ChatGPT could manage polypharmacy and make decisions on deprescribing 2 to 3 inappropriate medications based on a vignette [29]. Another study compared the answers from ChatGPT and clinical pharmacists to real clinical cases and clinical competency assessments [68]. However, these studies are different from this study, which used a systemic approach to polypharmacy (Multimedia Appendix 6). To the best of our knowledge, ChatGPT has not approached falls in older adults.

Future Directions

This study was a pilot with both promising findings and limitations. With the fast-growing use of ChatGPT and the new version of GPT-4o being released, the author suggests the following research directions.

A larger study using ChatGPT is needed to extend and confirm the findings of this study before adopting ChatGPT for geriatrics education and clinical practice. For example, more geriatrics knowledge questions and vignettes should be examined using GPT-4o. For example, the performance of ChatGPT on geriatrics certification exams should be tested.

Coinvestigators are needed to reduce biases. In particular, the correctness of the ChatGPT outputs should be evaluated and judged by a group of experts in geriatrics and integrated with the latest evidence, such as clinical practice guidelines, systematic reviews, and randomized controlled trials. For example, the ChatGPT outputs to a prompt should be rated from *strongly agree* to *strongly disagree* by a group of geriatrics experts in addition to determining their consistency with evidence-based clinical practice guidelines

Those who conduct ChatGPT research should receive training on prompt engineering. The author is doing this and feels that it is beneficial. This could reduce the variation in prompts.

The reliability of ChatGPT performance needs to be tested using different prompts to improve generalizability. The outcomes of the application of ChatGPT to geriatrics practice and education need to be further investigated.

Conclusions

This study suggests that ChatGPT could be a valuable assistant tool in geriatrics education and practice, helping health professional trainees and providers combat ageism, supplement geriatrics knowledge, and address common geriatric syndromes such as polypharmacy and falls. This study demonstrated that we could trust ChatGPT to be used in geriatrics practice and education by using 3 distinct approaches. One strength of this study is that Multimedia Appendices 5-7 provide the details of prompts to ChatGPT and the outputs generated by ChatGPT, which allows readers to apply a similar approach to their geriatrics education and practice studies. The findings of this study are promising but need more investigation before ChatGPT is widely adopted in geriatrics practice and education.

Data Availability

All data generated or analyzed during this study including all ChatGPT prompts and outputs are included in this published article and its supplementary information files on the journal website.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Geriatrics attitudes tests.
[DOCX File , 23 KB - formative_v9i1e63494_app1.docx]

Multimedia Appendix 2
University of California, Los Angeles, geriatrics knowledge test (questions 1-18).
[DOCX File , 25 KB - formative_v9i1e63494_app2.docx]

Multimedia Appendix 3

Comparison of geriatrics attitude and knowledge test performance between ChatGPT and trainees.
[DOCX File , 25 KB - formative_v9i1e63494_app3.docx]

Multimedia Appendix 4
Comparison of performance on the geriatrics attitude subscales by ChatGPT, trainees, and neurologists.
[DOCX File , 17 KB - formative_v9i1e63494_app4.docx]

Multimedia Appendix 5
Examples of ChatGPT outputs on geriatrics attitude and knowledge questions.
[DOCX File , 15 KB - formative_v9i1e63494_app5.docx]

Multimedia Appendix 6
ChatGPT output on the polypharmacy vignette about a woman aged 85 years.
[DOCX File , 17 KB - formative_v9i1e63494_app6.docx]

Multimedia Appendix 7
Fall risk identification and management in a woman aged 82 years.
[DOCX File , 20 KB - formative_v9i1e63494_app7.docx]

References

1. Number of ChatGPT users (Dec 2024). Exploding Topics. 2024 Dec 3. URL: https://explodingtopics.com/blog/chatgpt-users [accessed 2024-10-26]
2. PubMed homepage. PubMed. URL: https://pubmed.ncbi.nlm.nih.gov/ [accessed 2024-05-15]
3. Boscardin CK, Gin B, Golde PB, Hauer KE. ChatGPT and generative artificial intelligence for medical education: potential impact and opportunity. Acad Med 2023 Aug 31;99(1):22-27. [doi: 10.1097/acm.0000000000005439]
4. Preiksaitis C, Rose C. Opportunities, challenges, and future directions of generative artificial intelligence in medical education: scoping review. JMIR Med Educ 2023 Oct 20;9:e48785 [FREE Full text] [doi: 10.2196/48785] [Medline: 37862079]
5. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. Healthcare (Basel) 2023 Mar 19;11(6):887 [FREE Full text] [doi: 10.3390/healthcare11060887] [Medline: 36981544]
6. Thirunavukarasu AJ, Ting DS, Elangovan K, Gutierrez L, Tan TF, Ting DS. Large language models in medicine. Nat Med 2023 Aug 17;29(8):1930-1940. [doi: 10.1038/s41591-023-02448-8] [Medline: 37460753]
7. Meng X, Yan X, Zhang K, Liu D, Cui X, Yang Y, et al. The application of large language models in medicine: a scoping review. iScience 2024 May 17;27(5):109713 [FREE Full text] [doi: 10.1016/j.isci.2024.109713] [Medline: 38746668]
8. Liu J, Wang C, Liu S. Utility of ChatGPT in clinical practice. J Med Internet Res 2023 Jun 28;25:e48568 [FREE Full text] [doi: 10.2196/48568] [Medline: 37379067]
9. Andrew A. Potential applications and implications of large language models in primary care. Fam Med Community Health 2024 Jan 30;12(Suppl 1):e002602 [FREE Full text] [doi: 10.1136/fmch-2023-002602] [Medline: 38290759]
10. Li J, Dada A, Puladi B, Kleesiek J, Egger J. ChatGPT in healthcare: a taxonomy and systematic review. Comput Methods Programs Biomed 2024 Mar;245:108013 [FREE Full text] [doi: 10.1016/j.cmpb.2024.108013] [Medline: 38262126]
11. Rao SJ, Isath A, Krishnan P, Tangsrivimol JA, Virk HU, Wang Z, et al. ChatGPT: a conceptual review of applications and utility in the field of medicine. J Med Syst 2024 Jun 05;48(1):59. [doi: 10.1007/s10916-024-02075-x] [Medline: 38836893]
12. Momenaei B, Mansour HA, Kuriyan AE, Xu D, Sridhar J, Ting DS, et al. ChatGPT enters the room: what it means for patient counseling, physician education, academics, and disease management. Curr Opin Ophthalmol 2024 May 01;35(3):205-209. [doi: 10.1097/ICU.0000000000001036] [Medline: 38334288]
13. Cheng SW, Chang CW, Chang WJ, Wang HW, Liang CS, Kishimoto T, et al. The now and future of ChatGPT and GPT in psychiatry. Psychiatry Clin Neurosci 2023 Nov 11;77(11):592-596 [FREE Full text] [doi: 10.1111/pcn.13588] [Medline: 37612880]
14. Temperley HC, O'Sullivan NJ, Mac Curtain BM, Corr A, Meaney JF, Kelly ME, et al. Current applications and future potential of ChatGPT in radiology: a systematic review. J Med Imaging Radiat Oncol 2024 Apr;68(3):257-264. [doi: 10.1111/1754-9485.13621] [Medline: 38243605]
15. Lautrup AD, Hyrup T, Schneider-Kamp A, Dahl M, Lindholt JS, Schneider-Kamp P. Heart-to-heart with ChatGPT: the impact of patients consulting AI for cardiovascular health advice. Open Heart 2023 Nov;10(2):e002455 [FREE Full text] [doi: 10.1136/openhrt-2023-002455] [Medline: 37945282]
16. Mohammad B, Supti T, Alzubaidi M, Shah H, Alam T, Shah Z, et al. The pros and cons of using ChatGPT in medical education: a scoping review. Stud Health Technol Inform 2023 Jun 29;305:644-647. [doi: 10.3233/SHTI230580] [Medline: 37387114]

17. Duffourc M, Gerke S. Generative AI in health care and liability risks for physicians and safety concerns for patients. JAMA 2023 Jul 25;330(4):313-314. [doi: 10.1001/jama.2023.9630] [Medline: 37410497]
18. Kanter GP, Packel EA. Health care privacy risks of AI chatbots. JAMA 2023 Jul 25;330(4):311-312. [doi: 10.1001/jama.2023.9618] [Medline: 37410449]
19. Hanna JJ, Wakene AD, Lehmann CU, Medford RJ. Assessing racial and ethnic bias in text generation for healthcare-related tasks by ChatGPT. medRxiv. Preprint posted online on August 28, 2023 [FREE Full text] [doi: 10.1101/2023.08.28.23294730] [Medline: 37693388]
20. Amin KS, Forman HP, Davis MA. Even with ChatGPT, race matters. Clin Imaging 2024 May;109:110113. [doi: 10.1016/j.clinimag.2024.110113] [Medline: 38552383]
21. Erren TC, Lewis P, Shaw DM. Brave (in a) new world: an ethical perspective on chatbots for medical advice. Front Public Health 2023 Aug 17;11:1254334 [FREE Full text] [doi: 10.3389/fpubh.2023.1254334] [Medline: 37663854]
22. Korcok M. Medical schools face challenge of preparing physicians to care for fast-growing elderly population. JAMA 1985 Mar 01;253(9):1225-7, 1231. [Medline: 3968847]
23. Besdine R, Boult C, Brangman S, Coleman EA, Fried LP, Gerety M, et al. Caring for older Americans: the future of geriatric medicine. J Am Geriatr Soc 2005 Jun;53(6 Suppl):S245-S256. [doi: 10.1111/j.1532-5415.2005.53350.x] [Medline: 15963180]
24. Institute of Medicine, Committee on the Future Health Care Workforce for Older Americans, Board on Health Care Services. Retooling for an Aging America: Building the Health Care Workforce. Washington, DC: National Academies Press; 2008.
25. Dawson CM, Abiola AO, Sullivan AM, Schwartz AW, Members of the GERI Team Research Group. You can't be what you can't see: a systematic website review of Geriatrics Online-Visibility at US medical schools. J Am Geriatr Soc 2022 Oct;70(10):2996-3005 [FREE Full text] [doi: 10.1111/jgs.17997] [Medline: 36082829]
26. Gurwitz JH. The paradoxical decline of geriatric medicine as a profession. JAMA 2023 Aug 22;330(8):693-694. [doi: 10.1001/jama.2023.11110] [Medline: 37540519]
27. Fear K, Gleber C. Shaping the future of older adult care: ChatGPT, advanced AI, and the transformation of clinical practice. JMIR Aging 2023 Sep 13;6:e51776 [FREE Full text] [doi: 10.2196/51776] [Medline: 37703085]
28. Haque N. Artificial intelligence and geriatric medicine: new possibilities and consequences. J Am Geriatr Soc 2023 Jun 17;71(6):2028-2031. [doi: 10.1111/jgs.18334] [Medline: 36930059]
29. Rao A, Kim J, Lie W, Pang M, Fuh L, Dreyer KJ, et al. Proactive polypharmacy management using large language models: opportunities to enhance geriatric care. J Med Syst 2024 Apr 18;48(1):41. [doi: 10.1007/s10916-024-02058-y] [Medline: 38632172]
30. Rosselló-Jiménez D, Docampo S, Collado Y, Cuadra-Llopart L, Riba F, Llonch-Masriera M. Geriatrics and artificial intelligence in Spain (Ger-IA project): talking to ChatGPT, a nationwide survey. Eur Geriatr Med 2024 Aug;15(4):1129-1136. [doi: 10.1007/s41999-024-00970-7] [Medline: 38615289]
31. Nicholson K, Liu W, Fitzpatrick D, Hardacre KA, Roberts S, Salerno J, et al. Prevalence of multimorbidity and polypharmacy among adults and older adults: a systematic review. Lancet Healthy Longev 2024 Apr;5(4):e287-e296 [FREE Full text] [doi: 10.1016/S2666-7568(24)00007-2] [Medline: 38452787]
32. Hung A, Kim YH, Pavon JM. Deprescribing in older adults with polypharmacy. BMJ 2024 May 07;385:e074892. [doi: 10.1136/bmj-2023-074892] [Medline: 38719530]
33. Ng R, Chow TY. Aging narratives over 210 years (1810-2019). J Gerontol B Psychol Sci Soc Sci 2021 Oct 30;76(9):1799-1807 [FREE Full text] [doi: 10.1093/geronb/gbaa222] [Medline: 33300996]
34. Global report on ageism. World Health Organization. 2021 Mar 18. URL: https://www.who.int/teams/social-determinants-of-health/demographic-change-and-healthy-ageing/combatting-ageism/global-report-on-ageism [accessed 2024-12-12]
35. Hu RX, Luo M, Zhang A, Li LW. Associations of ageism and health: a systematic review of quantitative observational studies. Res Aging 2021 Aug;43(7-8):311-322. [doi: 10.1177/0164027520980130] [Medline: 33317402]
36. Allen JO, Solway E, Kirch M, Singer D, Kullgren JT, Moïse V, et al. Experiences of everyday ageism and the health of older US adults. JAMA Netw Open 2022 Jun 01;5(6):e2217240 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.17240] [Medline: 35704314]
37. Chang ES, Kannoth S, Levy S, Wang SY, Lee JE, Levy BR. Global reach of ageism on older persons' health: a systematic review. PLoS One 2020;15(1):e0220857 [FREE Full text] [doi: 10.1371/journal.pone.0220857] [Medline: 31940338]
38. Araújo PO, Soares IM, Vale PR, Sousa AR, Aparicio EC, Carvalho ES. Ageism directed to older adults in health services: a scoping review. Rev Lat Am Enfermagem 2023;31:e4019 [FREE Full text] [doi: 10.1590/1518-8345.6727.4019] [Medline: 37820219]
39. Amundsen D. A critical gerontological framing analysis of persistent ageism in NZ online news media: don't call us "elderly"!. J Aging Stud 2022 Jun;61:101009 [FREE Full text] [doi: 10.1016/j.jaging.2022.101009] [Medline: 35654544]
40. Murphy E, Fallon A, Dukelow T, O'Neill D. Don't call me elderly: a review of medical journals' use of ageist literature. Eur Geriatr Med 2022 Aug;13(4):1007-1009. [doi: 10.1007/s41999-022-00650-4] [Medline: 35581468]
41. Haase KR, Sattar S, Pilleron S, Lambrechts Y, Hannan M, Navarrete E, et al. A scoping review of ageism towards older adults in cancer care. J Geriatr Oncol 2023 Jan;14(1):101385. [doi: 10.1016/j.jgo.2022.09.014] [Medline: 36244925]

42. Bowling A. Ageism in cardiology. BMJ 1999 Nov 20;319(7221):1353-1355 [FREE Full text] [doi: 10.1136/bmj.319.7221.1353] [Medline: 10567146]
43. Silva MF, Silva DS, Bacurau AG, Francisco PM, Assumpção D, Neri AL, et al. Ageism against older adults in the context of the COVID-19 pandemic: an integrative review. Rev Saude Publica 2021 Apr 05;55:4 [FREE Full text] [doi: 10.11606/s1518-8787.2021055003082] [Medline: 33886953]
44. Altın Z, Buran F. Attitudes of health professionals toward elderly patients during the COVID-19 pandemic. Aging Clin Exp Res 2022 Oct;34(10):2567-2576 [FREE Full text] [doi: 10.1007/s40520-022-02209-6] [Medline: 35986878]
45. Martínez-Arnau FM, López-Hernández L, Castellano-Rioja E, Botella-Navas M, Pérez-Ros P. Interventions to improve attitudes toward older people in undergraduate health and social sciences students. A systematic review and meta-analysis. Nurse Educ Today 2022 Mar;110:105269 [FREE Full text] [doi: 10.1016/j.nedt.2022.105269] [Medline: 35063781]
46. Apriceno M, Levy SR. Systematic review and meta-analyses of effective programs for reducing ageism toward older adults. J Appl Gerontol 2023 Jun 17;42(6):1356-1375 [FREE Full text] [doi: 10.1177/07334648231165266] [Medline: 37069824]
47. Oca M, Meller L, Wilson K, Parikh AO, McCoy A, Chang J, et al. Bias and inaccuracy in AI chatbot ophthalmologist recommendations. Cureus 2023 Sep;15(9):e45911 [FREE Full text] [doi: 10.7759/cureus.45911] [Medline: 37885556]
48. Kishimoto M, Nagoshi M, Williams S, Masaki KH, Blanchette PL. Knowledge and attitudes about geriatrics of medical students, internal medicine residents, and geriatric medicine fellows. J Am Geriatr Soc 2005 Jan 22;53(1):99-102. [doi: 10.1111/j.1532-5415.2005.53018.x] [Medline: 15667384]
49. Reuben DB, Lee M, Davis JWJ, Eslami MS, Osterweil DG, Melchiore S, et al. Development and validation of a geriatrics attitudes scale for primary care residents. J Am Geriatr Soc 1998 Nov 27;46(11):1425-1430. [doi: 10.1111/j.1532-5415.1998.tb06012.x] [Medline: 9809767]
50. Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, et al. How does ChatGPT perform on the united states medical licensing examination (USMLE)? The implications of large language models for medical education and knowledge assessment. JMIR Med Educ 2023 Feb 08;9:e45312 [FREE Full text] [doi: 10.2196/45312] [Medline: 36753318]
51. Yaneva V, Baldwin P, Jurich DP, Swygert K, Clauser BE. Examining ChatGPT performance on USMLE sample items and implications for assessment. Acad Med 2024 Feb 01;99(2):192-197. [doi: 10.1097/ACM.0000000000005549] [Medline: 37934828]
52. Levin G, Horesh N, Brezinov Y, Meyer R. Performance of ChatGPT in medical examinations: a systematic review and a meta-analysis. BJOG 2024 Feb;131(3):378-380. [doi: 10.1111/1471-0528.17641] [Medline: 37604703]
53. Kung JE, Marshall C, Gauthier C, Gonzalez TA, Jackson JB3. Evaluating ChatGPT performance on the orthopaedic in-training examination. JB JS Open Access 2023;8(3):e23.00056 [FREE Full text] [doi: 10.2106/JBJS.OA.23.00056] [Medline: 37693092]
54. Passby L, Jenko N, Wernham A. Performance of ChatGPT on specialty certificate examination in dermatology multiple-choice questions. Clin Exp Dermatol 2024 Jun 25;49(7):722-727. [doi: 10.1093/ced/llad197] [Medline: 37264670]
55. Ali R, Tang OY, Connolly ID, Zadnik Sullivan PL, Shin JH, Fridley JS, et al. Performance of ChatGPT and GPT-4 on neurosurgery written board examinations. Neurosurgery 2023 Dec 01;93(6):1353-1365. [doi: 10.1227/neu.0000000000002632] [Medline: 37581444]
56. Long C, Lowe K, Zhang J, Santos AD, Alanazi A, O'Brien D, et al. A novel evaluation model for assessing ChatGPT on otolaryngology-head and neck surgery certification examinations: performance study. JMIR Med Educ 2024 Jan 16;10:e49970 [FREE Full text] [doi: 10.2196/49970] [Medline: 38227351]
57. Reuben DB, Lee M, Davis JW, Eslami MS, Osterweil DG, Melchiore S, et al. Development and validation of a geriatrics knowledge test for primary care residents. J Gen Intern Med 1997 Jul;12(7):450-452 [FREE Full text] [doi: 10.1046/j.1525-1497.1997.00078.x] [Medline: 9229284]
58. Rao A, Pang M, Kim J, Kamineni M, Lie W, Prasad AK, et al. Assessing the utility of ChatGPT throughout the entire clinical workflow. medRxiv. Preprint posted online on February 26, 2023 [FREE Full text] [doi: 10.1101/2023.02.21.23285886] [Medline: 36865204]
59. Qu RW, Qureshi U, Petersen G, Lee SC. Diagnostic and management applications of ChatGPT in structured otolaryngology clinical scenarios. OTO Open 2023;7(3):e67 [FREE Full text] [doi: 10.1002/oto2.67] [Medline: 37614494]
60. Franco D'Souza R, Amanullah S, Mathew M, Surapaneni KM. Appraising the performance of ChatGPT in psychiatry using 100 clinical case vignettes. Asian J Psychiatr 2023 Nov;89:103770. [doi: 10.1016/j.ajp.2023.103770] [Medline: 37812998]
61. Goodman RS, Patrinely JR, Stone CAJ, Zimmerman E, Donald RR, Chang SS, et al. Accuracy and reliability of chatbot responses to physician questions. JAMA Netw Open 2023 Oct 02;6(10):e2336483 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.36483] [Medline: 37782499]
62. Kassab J, El Dahdah J, Chedid El Helou M, Layoun H, Sarraju A, Laffin LJ, et al. Assessing the accuracy of an online chat-based artificial intelligence model in providing recommendations on hypertension management in accordance with the 2017 American College of Cardiology/American Heart Association and 2018 European Society of Cardiology/European Society of Hypertension Guidelines. Hypertension 2023 Jul;80(7). [doi: 10.1161/hypertensionaha.123.21183]
63. Rizwan A, Sadiq T. The use of AI in diagnosing diseases and providing management plans: a consultation on cardiovascular disorders with ChatGPT. Cureus 2023 Aug;15(8):e43106 [FREE Full text] [doi: 10.7759/cureus.43106] [Medline: 37692649]

64. Zhang L, Tashiro S, Mukaino M, Yamada S. Use of artificial intelligence large language models as a clinical tool in rehabilitation medicine: a comparative test case. J Rehabil Med 2023 Sep 11;55:jrm13373 [FREE Full text] [doi: 10.2340/jrm.v55.13373] [Medline: 37691497]
65. Juhi A, Pipil N, Santra S, Mondal S, Behera JK, Mondal H. The capability of ChatGPT in predicting and explaining common drug-drug interactions. Cureus 2023 Mar;15(3):e36272 [FREE Full text] [doi: 10.7759/cureus.36272] [Medline: 37073184]
66. Fournier A, Fallet C, Sadeghipour F, Perrottet N. Assessing the applicability and appropriateness of ChatGPT in answering clinical pharmacy questions. Ann Pharm Fr 2024 May;82(3):507-513. [doi: 10.1016/j.pharma.2023.11.001] [Medline: 37992892]
67. Al-Dujaili Z, Omari S, Pillai J, Al Faraj A. Assessing the accuracy and consistency of ChatGPT in clinical pharmacy management: a preliminary analysis with clinical pharmacy experts worldwide. Res Social Adm Pharm 2023 Dec;19(12):1590-1594. [doi: 10.1016/j.sapharm.2023.08.012] [Medline: 37696742]
68. Huang X, Estau D, Liu X, Yu Y, Qin J, Li Z. Evaluating the performance of ChatGPT in clinical pharmacy: a comparative study of ChatGPT and clinical pharmacists. Br J Clin Pharmacol 2024 Jan;90(1):232-238. [doi: 10.1111/bcp.15896] [Medline: 37626010]
69. Choi W. Assessment of the capacity of ChatGPT as a self-learning tool in medical pharmacology: a study using MCQs. BMC Med Educ 2023 Nov 13;23(1):864 [FREE Full text] [doi: 10.1186/s12909-023-04832-x] [Medline: 37957666]
70. Cheng HY, Bradley E. Improving identification of potentially inappropriate medications among older patients: a preliminary report of a single-blind, controlled trial teaching fourth year medical students value- and preference-based prescribing/de-prescribing. Med Sci Educ 2016 Jun 28;26:431-435. [doi: 10.1007/s40670-016-0283-2]
71. Cheng H. Falls and fall prevention. MedEdPORTAL 2014 Sep 26:9910. [doi: 10.15766/mep_2374-8265.9910]
72. Alqahtani R, Almuhaidib S, Jradi H. A cross-sectional study: exploring knowledge and attitude of medical and nursing students to Care for Elders in the future. BMC Geriatr 2022 Nov 14;22(1):856 [FREE Full text] [doi: 10.1186/s12877-022-03551-0] [Medline: 36376782]
73. Hogan TM, Chan SB, Hansoti B. Multidimensional attitudes of emergency medicine residents toward older adults. West J Emerg Med 2014 Jul;15(4):511-517 [FREE Full text] [doi: 10.5811/westjem.2014.2.19937] [Medline: 25035760]
74. Seferoğlu M, Yıldız D, Pekel NB, Güneş A, Yıldız A, Tufan F. Attitudes of neurology specialists toward older adults. Aging Clin Exp Res 2017 Aug;29(4):787-792. [doi: 10.1007/s40520-016-0606-6] [Medline: 27456680]
75. O'Mahony D, Cherubini A, Guiteras AR, Denkinger M, Beuscart JB, Onder G, et al. STOPP/START criteria for potentially inappropriate prescribing in older people: version 3. Eur Geriatr Med 2023 Aug;14(4):625-632 [FREE Full text] [doi: 10.1007/s41999-023-00777-y] [Medline: 37256475]
76. Wilson MA, Kurrle S, Wilson I. Medical student attitudes towards older people: a critical review of quantitative measures. BMC Res Notes 2018 Jan 24;11(1):71 [FREE Full text] [doi: 10.1186/s13104-018-3186-z] [Medline: 29361969]
77. By the 2023 American Geriatrics Society Beers Criteria® Update Expert Panel. American Geriatrics Society 2023 updated AGS Beers Criteria® for potentially inappropriate medication use in older adults. J Am Geriatr Soc 2023 Jul 04;71(7):2052-2081. [doi: 10.1111/jgs.18372] [Medline: 37139824]
78. Hilmer SN, Mager DE, Simonsick EM, Cao Y, Ling SM, Windham BG, et al. A drug burden index to define the functional burden of medications in older people. Arch Intern Med 2007 Apr 23;167(8):781-787. [doi: 10.1001/archinte.167.8.781] [Medline: 17452540]
79. Samsa GP, Hanlon JT, Schmader KE, Weinberger M, Clipp EC, Uttech KM, et al. A summated score for the medication appropriateness index: development and assessment of clinimetric properties including content validity. J Clin Epidemiol 1994 Aug;47(8):891-896. [doi: 10.1016/0895-4356(94)90192-9] [Medline: 7730892]
80. Pravodelov V. Thoughtful prescribing and deprescribing. Med Clin North Am 2020 Sep;104(5):751-765. [doi: 10.1016/j.mcna.2020.06.001] [Medline: 32773043]
81. Ramaswamy R, Maio V, Diamond JJ, Talati AR, Hartmann CW, Arenson C, et al. Potentially inappropriate prescribing in elderly: assessing doctor knowledge, confidence and barriers. J Eval Clin Pract 2011 Dec;17(6):1153-1159. [doi: 10.1111/j.1365-2753.2010.01494.x] [Medline: 20630004]
82. Abd Wahab MS. The relevance of educating doctors, pharmacists and older patients about potentially inappropriate medications. Int J Clin Pharm 2015 Dec;37(6):971-974. [doi: 10.1007/s11096-015-0203-3] [Medline: 26415737]
83. Rathore A, Sharma R, Bansal P, Chhabra M, Arora M. Knowledge, attitude, and practice of medical interns and postgraduate residents on American Geriatric Society updated Beers criteria. J Educ Health Promot 2023;12:1 [FREE Full text] [doi: 10.4103/jehp.jehp_769_22] [Medline: 37034871]
84. Vandergrift JL, Weng W, Leff B, Gray BM. Geriatricians, general internists, and potentially inappropriate medications for a national sample of older adults. J Am Geriatr Soc 2024 Jan;72(1):37-47. [doi: 10.1111/jgs.18489] [Medline: 37350649]
85. Moreland B, Kakara R, Henry A. Trends in nonfatal falls and fall-related injuries among adults aged ≥65 years - United States, 2012-2018. MMWR Morb Mortal Wkly Rep 2020 Jul 10;69(27):875-881 [FREE Full text] [doi: 10.15585/mmwr.mm6927a5] [Medline: 32644982]
86. Xiong W, Wang D, Ren W, Liu X, Wen R, Luo Y. The global prevalence of and risk factors for fear of falling among older adults: a systematic review and meta-analysis. BMC Geriatr 2024 Apr 05;24(1):321 [FREE Full text] [doi: 10.1186/s12877-024-04882-w] [Medline: 38580924]

87. Rausch C, van Zon SK, Liang Y, Laflamme L, Möller J, de Rooij SE, et al. Geriatric syndromes and incident chronic health conditions among 9094 older community-dwellers: findings from the lifelines cohort study. J Am Med Dir Assoc 2022 Jan;23(1):54-9.e2 [FREE Full text] [doi: 10.1016/j.jamda.2021.02.030] [Medline: 33798484]
88. US Preventive Services Task Force, Nicholson WK, Silverstein M, Wong JB, Barry MJ, Chelmow D, et al. Interventions to prevent falls in community-dwelling older adults: US Preventive Services Task Force recommendation statement. JAMA 2024 Jul 02;332(1):51-57. [doi: 10.1001/jama.2024.8481] [Medline: 38833246]
89. Clinical resources. Centers for Disease Control and Prevention STEADI. 2024 Apr 22. URL: https://www.cdc.gov/steadi/hcp/clinical-resources/index.html [accessed 2024-06-25]
90. Xu Q, Ou X, Li J. The risk of falls among the aging population: a systematic review and meta-analysis. Front Public Health 2022;10:902599 [FREE Full text] [doi: 10.3389/fpubh.2022.902599] [Medline: 36324472]
91. 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]
92. Smith ML, Stevens JA, Ehrenreich H, Wilson AD, Schuster RJ, Cherry CO, et al. Healthcare providers' perceptions and self-reported fall prevention practices: findings from a large New York health system. Front Public Health 2015;3:17 [FREE Full text] [doi: 10.3389/fpubh.2015.00017] [Medline: 25964942]
93. Howland J, Hackman H, Taylor A, O'Hara K, Liu J, Brusch J. Older adult fall prevention practices among primary care providers at accountable care organizations: a pilot study. PLoS One 2018;13(10):e0205279 [FREE Full text] [doi: 10.1371/journal.pone.0205279] [Medline: 30307974]
94. Ortmann N, Haddad YK, Beck L. Special report from the CDC: provider knowledge and practices around driving safety and fall prevention screening and recommendations for their older adult patients, DocStyles 2019. J Safety Res 2023 Sep;86:401-408. [doi: 10.1016/j.jsr.2023.06.011] [Medline: 37718068]
95. Davenport K, Cameron A, Samson M, Sri-On J, Liu SW. Fall prevention knowledge, attitudes, and behaviors: a survey of emergency providers. West J Emerg Med 2020 Jul 10;21(4):826-830 [FREE Full text] [doi: 10.5811/westjem.2020.4.43387] [Medline: 32726252]
96. Meekes WM, Leemrijse CJ, Weesie YM, van de Goor IA, Donker GA, Korevaar JC. Falls prevention at GP practices: a description of daily practice. BMC Fam Pract 2021 Sep 21;22(1):190 [FREE Full text] [doi: 10.1186/s12875-021-01540-7] [Medline: 34548022]
97. Siontis KC, Attia ZI, Asirvatham SJ, Friedman PA. ChatGPT hallucinating: can it get any more humanlike? Eur Heart J 2024 Feb 01;45(5):321-323. [doi: 10.1093/eurheartj/ehad766] [Medline: 38088452]
98. Hatem R, Simmons B, Thornton JE. A call to address AI "hallucinations" and how healthcare professionals can mitigate their risks. Cureus 2023 Sep;15(9):e44720 [FREE Full text] [doi: 10.7759/cureus.44720] [Medline: 37809168]
99. Sallam M, Barakat M, Sallam M. A preliminary checklist (METRICS) to standardize the design and reporting of studies on generative artificial intelligence-based models in health care education and practice: development study involving a literature review. Interact J Med Res 2024 Feb 15;13:e54704 [FREE Full text] [doi: 10.2196/54704] [Medline: 38276872]

Abbreviations

AI: artificial intelligence
CDC: Centers for Disease Control and Prevention
METRICS: Model, Evaluation, Timing, Range and Randomization, Individual Factors, Counts, and Specificity of Prompts and Language
PIM: potentially inappropriate medication
UCLA: University of California, Los Angeles
USMLE: United States Medical Licensing Examination

Edited by A Mavragani; submitted 30.06.24; peer-reviewed by J Mira, P Heyn; comments to author 17.10.24; revised version received 26.10.24; accepted 17.11.24; published 03.01.25.

Please cite as:
Cheng HY
ChatGPT's Attitude, Knowledge, and Clinical Application in Geriatrics Practice and Education: Exploratory Observational Study
JMIR Form Res 2025;9:e63494
URL: https://formative.jmir.org/2025/1/e63494
doi:10.2196/63494
PMID:39752214

Original Paper

A Digital Parenting Intervention With Intimate Partner Violence Prevention Content: Quantitative Pre-Post Pilot Study

Moa Schafer[1], DPhil; Jamie Lachman[1,2], DPhil; Paula Zinser[1], MSc; Francisco Antonio Calderón Alfaro[1], MSc; Qing Han[1], PhD; Chiara Facciola[3], PhD; Lily Clements[3], PhD; Frances Gardner[1], DPhil; Genevieve Haupt Ronnie[2], MA; Ross Sheil[4], BSc

[1]Centre for Evidence Based Intervention, Department of Social Policy and Intervention, University of Oxford, Oxford, United Kingdom

[2]Centre for Social Science Research, University of Cape Town, Cape Town, South Africa

[3]Innovations in Development, Education, and the Mathematical Sciences (IDEMS) International, Reading, United Kingdom

[4]UNICEF Jamaica, Kingston, Jamaica

Corresponding Author:
Moa Schafer, DPhil
Centre for Evidence Based Intervention
Department of Social Policy and Intervention
University of Oxford
Barnett House
32-37 Wellington Square
Oxford, OX1 2ER
United Kingdom
Phone: 44 01865 2703
Email: moa.schafer@spi.ox.ac.uk

Abstract

Background: Intimate partner violence (IPV) and violence against children are global issues with severe consequences. Intersections shared by the 2 forms of violence have led to calls for joint programming efforts to prevent both IPV and violence against children. Parenting programs have been identified as a key entry point for addressing multiple forms of family violence. Building on the IPV prevention material that has been integrated into the parenting program ParentText, a digital parenting chatbot, this pilot study seeks to explore parents' engagement with the IPV prevention content in ParentText and explore preliminary changes in IPV.

Objective: This study aimed to assess parents' and caregivers' level of engagement with the IPV prevention material in the ParentText chatbot and explore preliminary changes in experiences and perpetration of IPV, attitudes toward IPV, and gender-equitable behaviors following the intervention.

Methods: Caregivers of children aged between 0 and 18 years were recruited through convenience sampling by research assistants in Cape Town, South Africa, and by UNICEF (United Nations Children's Fund) Jamaica staff in 3 parishes of Jamaica. Quantitative data from women in Jamaica (n=28) and South Africa (n=19) and men in South Africa (n=21) were collected electronically via weblinks sent to caregivers' phones using Open Data Kit. The primary outcome was IPV experience (women) and perpetration (men), with secondary outcomes including gender-equitable behaviors and attitudes toward IPV. Descriptive statistics were used to report sociodemographic characteristics and engagement outcomes. Chi-square tests and 2-tailed paired dependent-sample *t* tests were used to investigate potential changes in IPV outcomes between pretest and posttest.

Results: The average daily interaction rate with the program was 0.57 and 0.59 interactions per day for women and men in South Africa, and 0.21 for women in Jamaica. The rate of completion of at least 1 IPV prevention topic was 25% (5/20) for women and 5% (1/20) for men in South Africa, and 21% (6/28) for women in Jamaica. Exploratory analyses indicated significant pre-post reductions in overall IPV experience among women in South Africa (*P*=.01) and Jamaica (*P*=.01) and in men's overall harmful IPV attitudes (*P*=.01) and increases in men's overall gender-equitable behaviors (*P*=.02) in South Africa.

Conclusions: To the best of our knowledge, this is the first pilot study to investigate user engagement with and indicative outcomes of a digital parenting intervention with integrated IPV prevention content. Study findings provide valuable insights into user interactions with the chatbot and shed light on challenges related to low levels of chatbot engagement. Indicative results

suggest promising yet modest reductions in IPV and improvements in attitudes after the program. Further research using a randomized controlled trial is warranted to establish causality.

(JMIR Form Res 2025;9:e58611) doi:10.2196/58611

KEYWORDS

intimate partner violence; SMS text messaging; chatbot; user engagement; parenting; violence; mobile phone

Introduction

Background

Violence against women (VAW) and violence against children (VAC) are global issues with severe, long-lasting consequences, which affect individuals and communities worldwide. Global reports have revealed alarmingly high rates of both forms of violence [1,2]. Prevalence estimates of intimate partner violence (IPV), which is the most common form of VAW, have found that >27% of ever-partnered women aged >15 years have experienced physical IPV, sexual IPV, or both at least once in their life [1]. Reports of VAC are also of great concern, with systematic review findings suggesting that >50% of all children worldwide, that is, >1 billion children globally, have experienced past-year violence [3,4]. Research findings also indicate that prevalence rates are higher for both forms of violence in low- and middle-income countries [5,6].

A growing body of evidence shows that IPV and VAC frequently cooccur in the same families [7]. In recent years, the intersections between IPV and VAC have received increasing attention, with a growing demand for prevention efforts that target both forms of violence concurrently [8]. Intersections shared by both IPV and VAC include common risk factors, such as acceptability of family violence and parental history of physical abuse [9,10], similar short- and long-term consequences, including mental and physical health problems [11], and various intergenerational effects [12]. In the past, despite their overlaps, the fields of IPV and VAC prevention research have predominantly been separated [8]. Therefore, initiatives that seek to address these common risk factors may help shift violence prevention efforts from a siloed approach to one that targets multiple types of violence simultaneously. By targeting multiple and intergenerational forms of violence, this strategy is likely to be more sustainable long-term, helping to reduce costs and increasing opportunities for scale-up and impact [13].

In the field of IPV prevention, researchers and practitioners are increasingly adopting what is known as gender-transformative approaches in programs that aim to address multiple forms of violence, that is, both VAW and VAC [14]. More specifically, gender-transformative approaches can be defined as those that aim to address harmful social and gender norms and are "designed specifically to encourage men and boys to adopt and enact gender-equitable, nonviolent attitudes and behaviors" [15]. Many programs that adopt this approach include activities and modules that focus on topics such as improving interpersonal skills, redefining responsibilities in the home, and changing perceptions around gender roles and violence [14]. Emerging evidence suggests that gender-transformative approaches offer promising results in preventing IPV perpetration, shifting restrictive gender norms, and promoting gender-equitable behaviors and attitudes [16-18]. Gender-transformative approaches are being adopted in family and sexual health interventions, further highlighting the shift toward a more holistic approach toward violence prevention and recognition of the intergenerational elements of family violence and shared risk factors [19,20].

Alongside gender-transformative interventions, in the field of VAC, parenting programs have been identified as a major strategy for reducing and preventing violence [21]. Findings from a recent systematic review of parenting programs for reducing child maltreatment in low- and middle-income countries identified several promising results across child, parent, and family outcomes, with meta-analyses results revealing reductions in child maltreatment and harsh parenting [22]. One program that has integrated both VAC and IPV prevention content includes the gender-transformative Bandebereho program in Rwanda, which engaged men and partners with the aim of encouraging healthier couple relationships and enhancing caregiving skills [23]. The program used fatherhood as an avenue to promote gender equality and foster positive shifts in men's relations with both their children and partners [18]. A randomized controlled trial (RCT) of the program found promising short- and long-term effects, with reductions in IPV and physical punishment of children at both 21 months posttest and at a 6-year follow-up [18].

Despite the promising potential of parenting programs to tackle multiple forms of family violence, scaling up parenting interventions remains a challenge [24]. To address barriers to scale, digital parenting interventions have been gaining momentum as a means to increase reach, engagement, and accessibility [25]. One such program, developed by Parenting for Lifelong Health [26,27], is a chatbot intervention called ParentText that provides caregivers with social learning–based parenting content that seeks to improve parent-child interaction and prevent child maltreatment [28]. Given the emerging evidence based on the potential to integrate parenting and gender-transformative approaches to prevent multiple forms of family violence, this study seeks to explore how users engage with a parenting chatbot that contains integrated IPV prevention material. Building on the development of IPV prevention material designed and integrated into the ParentText intervention in 2021 [28], the purpose of this pilot study was to examine user engagement with the IPV prevention content in the chatbot as well as explore preliminary changes in IPV.

Objectives

The overarching aim of this pilot study is 2-fold. First, it seeks to assess parents' and caregivers' level of engagement with the integrated IPV prevention material in the ParentText chatbot.

Second, it aims to conduct exploratory analyses of preliminary pre-post changes in experiences and perpetration of IPV and related outcomes, including attitudes toward IPV and gender-equitable behaviors following the intervention.

Consequently, this study seeks to answer the following research questions:

1. What is the level of user engagement with the IPV prevention content integrated into the parenting chatbot, ParentText, among parents and caregivers with low income in South Africa and Jamaica?
2. What, if any, are the preliminary changes in experiences and perpetration of IPV (primary outcomes) and related risk factors, including attitudes toward IPV and gender-equitable behaviors (secondary outcomes) among caregivers with low income in South Africa and Jamaica, after using the ParentText chatbot?

Methods

Study Setting

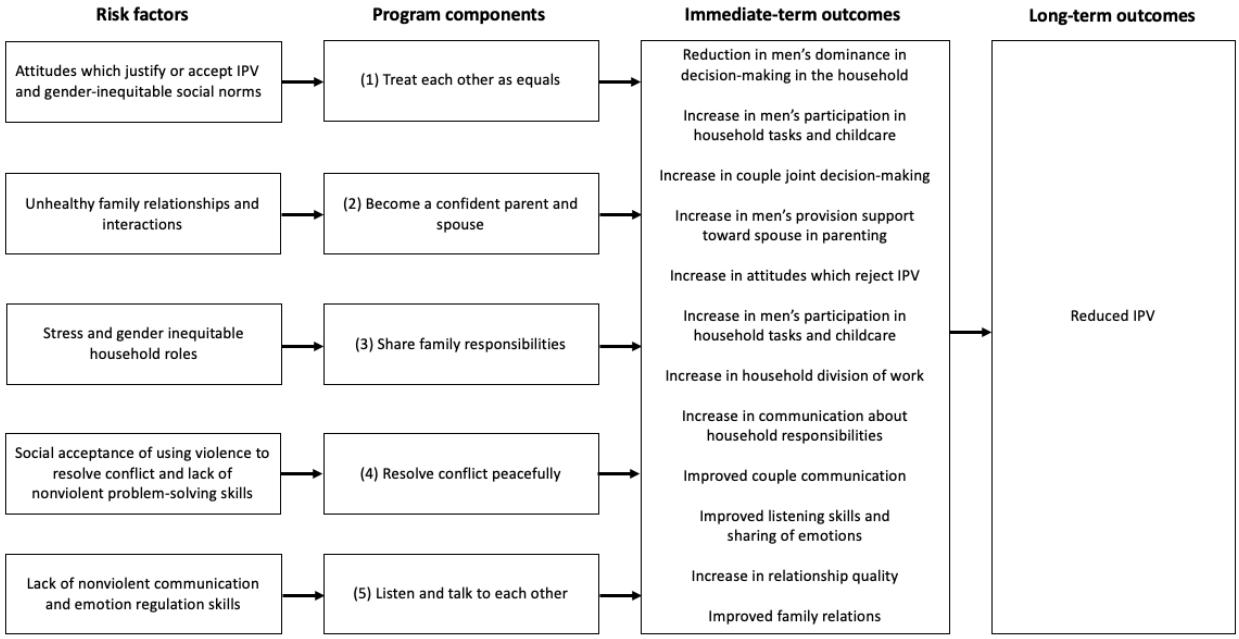
The study was conducted in 2 countries where IPV is particularly widespread: South Africa and Jamaica. Despite differing in their cultural contexts and geographical regions, South Africa and Jamaica share similar challenges surrounding high rates of IPV, making them comparable settings for exploring the integration of IPV prevention strategies in a parenting chatbot. South Africa has one of the highest rates of IPV globally, with studies revealing prevalence rates of 20% to 50% in which women reported having experienced IPV sometime during their life [29-31]. In Jamaica, rates of VAW are also alarming, with 24% of women having experienced physical violence, sexual violence, or both, from an intimate partner sometime in their life [32]. These high rates of violence underscore that urgent prevention efforts that seek to prevent and reduce violence are crucial. Moreover, research findings also indicate a high prevalence of patriarchal social norms and restrictive beliefs around women's roles in both Jamaica [33] and South Africa [34], further highlighting the harmful sociocultural norms prevalent in both contexts and the overlapping challenges that both countries face. In Jamaica, this study was conducted in the parishes: Kingston, St Catherine, and St Elizabeth; and in South Africa, this study was conducted in Cape Town.

Participants and Recruitment

In Cape Town, South Africa, participants were recruited through convenience sampling by local research assistants through the organization Clowns Without Borders South Africa. Recruitment mainly took place in neighborhoods, townships, and communities in urban Cape Town. In Jamaica, recruitment also took place through convenience sampling by members of UNICEF (United Nations Children's Fund) Jamaica staff members in partnership with the National Parenting Support Commission, an agency of the Jamaican Ministry of Education and Youth. Participants were recruited primarily at schools and at women's centers for adolescent mothers in urban and rural communities. While the aim was to recruit both men and women, due to resource constraints, in the end, only women were recruited in Jamaica. Furthermore, while originally, the study sought to only include parents aged >18 years, following an interest expressed by UNICEF Jamaica and the Jamaican National Parenting Support Commission to address challenges faced by adolescent parents in the country, additional procedures were put in place (refer to the Ethical Considerations section for more details) to include adolescent parents aged >16 years in Jamaica. Unfortunately, due to staff and resource limitations in Cape Town, it was only possible to recruit parents aged >18 years in South Africa.

Consequently, the inclusion criteria for participants in this study were as follows: (1) age >18 years in South Africa and >16 years in Jamaica (where adolescent parents aged >16 years were also recruited), (2) currently caring for a child aged between 0 and 17 years, (3) being in a relationship (defined as either having a partner or being married—refer to Multimedia Appendix 1 for further details), (4) having access to a phone that could receive messages on WhatsApp (Meta Platforms) or Telegram (Telegram Messenger Inc) and that could connect to 3G or the internet, (5) provided consent to participate in the study, and (6) being able to speak either English or isiXhosa (South Africa only). Exclusion criteria were participants who were not currently in a partnered relationship and participants who were not caregivers, parents, or currently caring for a child. As an initial pre-post evaluation, this study was not designed to detect significant intervention effects but rather to assess user engagement and conduct exploratory analyses of intervention outcomes [35]. Due to funding limitations, the sample size of the study was restricted to 86 participants, with some lost due to drop out (Figure 1). Nevertheless, the study used a G*Power 3 calculator with a sensitivity power analysis to calculate the effect size (Cohen *d*) needed to obtain a significant intervention effect. Input parameters included the use of 2-tailed paired *t* tests based on the study's primary outcomes. Assuming a type I error of $P<.05$, 80% power, this sample size was sufficiently powered to detect a moderately significant intervention effect of Cohen $d=0.67$.

Figure 1. Theory of change for the ParentText intimate partner violence (IPV) prevention content, including risk factors, program components, immediate-term outcomes, and long-term outcomes.



Intervention

ParentText is a chatbot intervention for parents and caregivers of children aged between 0 and 17 years [28]. The technical architecture is open source and available for developers through GitHub. The digital parenting intervention sends automated messages to users through social messaging platforms, such as WhatsApp, Telegram, and Facebook messenger (Meta Platforms), and is also available via SMS text messaging for individuals without smartphone access (for an example of program components and example messages, refer to Figures 1-3 and Multimedia Appendix 2). The parenting content of ParentText is derived from the in-person Parenting for Lifelong Health programs [26,27]. Enrolled users receive 23 days of daily ParentText messages personalized for their chosen child's developmental stage (0-23 months, 2-9 years, and 10-17 years). Content is provided using a range of formats, including text, image, video, and audio. Structured and sequential parenting content is derived from common elements of social learning theory-based programs along two main themes as follows: (1) positive relationship building and (2) limit setting and nonviolent discipline [36]. In addition, content specifically targeting positive partner relationships and IPV prevention for users in partnered relationships was integrated into the chatbot by drawing on gender-transformative material from a range of interventions found to reduce IPV and improve gender-equitable attitudes [23,37,38]. It is the engagement with this content that is the primary focus of this study. The IPV prevention content in ParentText is divided into 5 topics (refer to program components in Figure 1). The proposed causal pathway between each IPV prevention topic, identified risk factor, and behavior change domain is illustrated in the theory of change in Figure 1. Schafer et al [28] and Multimedia Appendix 3 give further details on the development of the IPV prevention content and theory of change. Additional information and examples on how parents can interact with the chatbot, along with examples of the sequencing of the ParentText messages, are available in Figures 2 and 3.

Figure 2. Flow of the intimate partner violence (IPV) prevention content text messages in ParentText, which illustrates the flow of text messages that users in ParentText receive, including introduction text messages and text messages providing options to view further IPV prevention content.

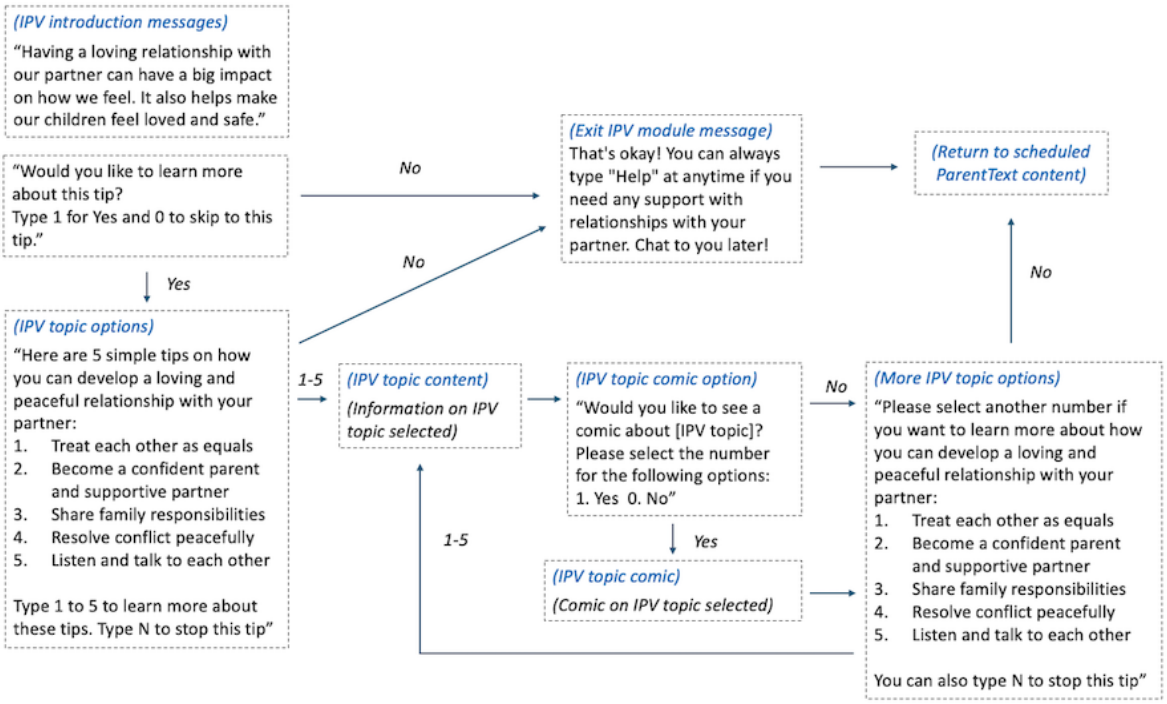
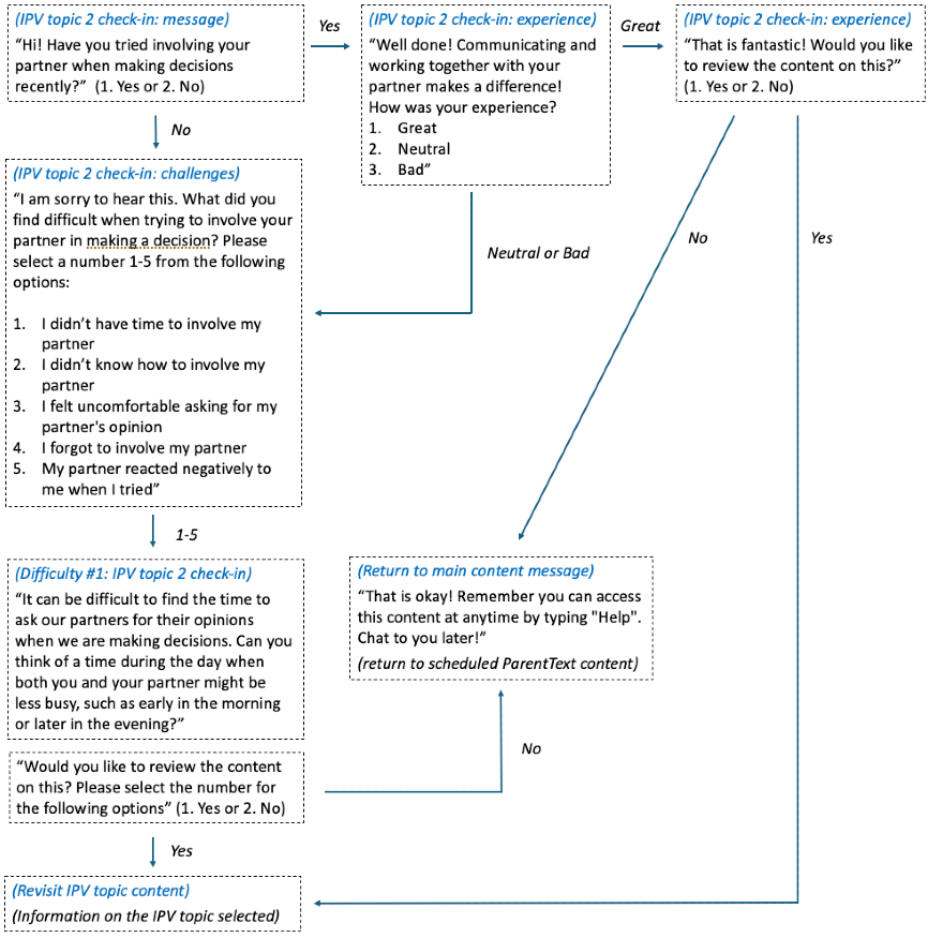


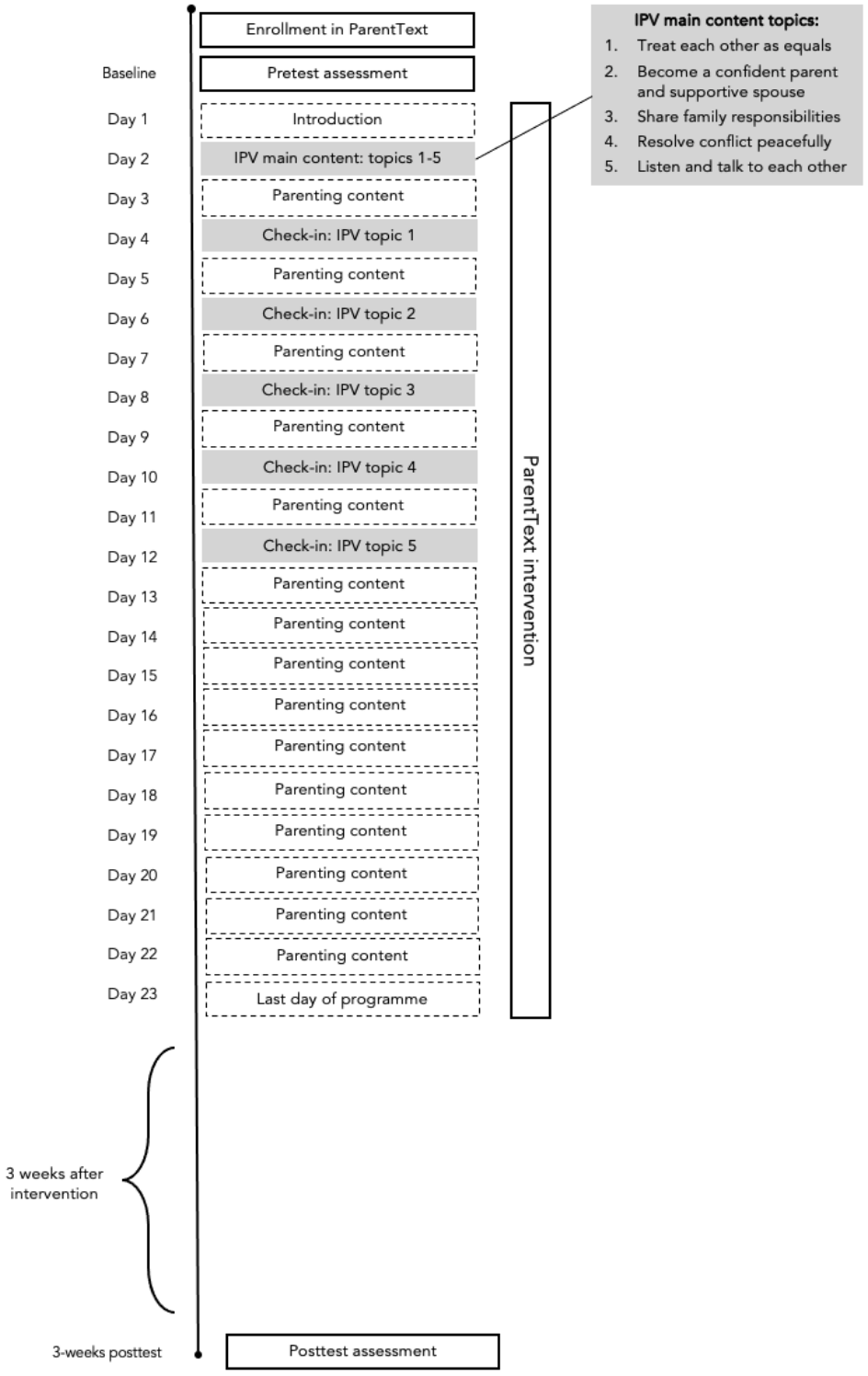
Figure 3. Flow of the check-in message on the intimate partner violence (IPV) topic 2 "joint decision-making," which illustrates the flow of the check-in text message that users in ParentText receive to provide an opportunity to reflect on the topic of "joint decision-making.".



Design and Procedures

This study used a repeated-measures, single-arm design. Pre and posttest assessments were conducted at baseline (preintervention) and at 6 weeks after baseline (ie, 3 weeks after the end of the intervention) using the application Open Data Kit (ODK; refer to the Measures section and Multimedia Appendix 1 for more details). The posttest was planned as a 1-month posttest (1 month after the end of intervention); however, due to fieldwork challenges that arose and time constraints, it needed to be changed to 6 weeks after baseline (Figure 4).

Figure 4. Overview of the ParentText chatbot timeline, illustrating what content is delivered on what days of the program from day 1 to day 23, and demonstrating when the 3-week posttest assessment is delivered. IPV: intimate partner violence.



Reporting follows the CONSORT (Consolidated Standards of Reporting Trials) 2010 guidelines extension to randomized pilot and feasibility trials [39] (refer to Multimedia Appendix 4 for the CONSORT checklist). Recruitment, delivery, and

assessments in South Africa and Jamaica were conducted on a rolling basis from September 2022 to March 2023. Participants who enrolled in the intervention received data credit using a local service provider to ensure they would have access to 3G during the intervention. During recruitment, participants received a message that included a link to the program embedded within a WhatsApp or Telegram business account. After clicking the link, participants were directed to the ParentText chatbot account and asked to select their preferred language for the program. In keeping with recommendations from the formative research [28], the language was set to English in Jamaica, and the options were English and isiXhosa in South Africa. Participants were asked if they consented to participate in the intervention and evaluation and were provided with a link to a more detailed information sheet. Participants who did not consent exited the intervention. In alignment with guidelines for conducting statistical surveys on VAW [40], the survey included skip patterns, which determine the eligibility of specific questions that respondents are asked [40,41]. For example, in this study, men who responded that their partner was taking part in the intervention were not asked the IPV items in the IPV assessment and were instead only asked the questions on attitudes toward IPV and gender roles and on gender-equitable behaviors. Further information about this is provided in the Ethical Considerations section.

At baseline, participants were invited to complete the IPV assessment, which was optional, on an external server via Oxford's Linux virtual machine using ODK to ensure responses were not saved on their mobile devices. Participants provided their phone number when completing the IPV survey for their responses to be recorded and linked with a unique user ID. These data were then deidentified, exported as CSV files, and uploaded to a secure server in Oxford, which was only accessible by the research team and password protected. All personal identifying data were deleted once endline data collection was completed. Participants were invited to respond to the IPV assessment at 3 weeks postintervention via the phone number they provided when completing the IPV assessment at baseline.

Measures

The user engagement, demographic, primary outcomes, and secondary outcomes are provided in Table 1.

Table 1. Outcome measurements in the study, including user engagement, demographics, background variables, and primary and secondary outcomes, with descriptions of the study measurement used and the number of items used for each outcome.

Outcomes	Study measurement	Items, n
User engagement outcomes		
Number of days in the ParentText program	Measured by the final day the participant used the chatbot subtracted by the first day participant used the chatbot (including inactive days)	1
Number of active days the user actively interacted with the chatbot	Number of days participants actively interacted with ParentText as measured by chatbot interactions (excluding inactive days)	1
Overall chatbot interaction rate	The number of interactions logged per day divided by the total number of days participants actively interacted with ParentText	1
Participant engagement with each IPV[a] topic	The number of participants who viewed each IPV topic, in line with existing digital parenting intervention research where module viewing is frequently used as a proxy measure for the level of participant engagement [42,43]	5
Completion rate of IPV topics	Number of IPV topics completed out of the total 5 topics (measured by the number of IPV topics participants viewed that were recorded via RapidPro, UNICEF's open source framework developed to send and receive data using mobile phones [44])	5
Overall completion rate of ParentText parenting modules	Number of parenting modules completed (measured by the number of modules viewed that were recorded via the RapidPro (UNICEF) software)	1
Primary behavioral outcomes		
Women's experiences of IPV	Items adapted from the WHO[b] VAWI[c] and the core questionnaire in the WHO-MCS[d] [45]: physical (1 item), psychological (1 item), sexual (1 item), coercion (1 item), and economic abuse (1 item). Participants asked to give a frequency score on a scale of 0 to 8 or more times. The original VAWI tool measures past-year experiences of IPV [45]; however, a 1-year reporting time frame surpassed the duration of the intervention. The reporting timeframe and response codes were therefore adapted from the past-year (once, a few times, and many times) to the number of experiences of IPV in the past month (0, 1, 2, 3, 4, 5, 6, 7, and >8) to enhance the sensitivity of measuring changes in experiences of IPV. These adjustments are similar to those made to measure child maltreatment with the ISPCAN[e] Child Abuse Screening Tool (ICAST)-Trial measurement by Meinck et al [46]. Past-month time frame when measuring IPV has been used in other similar pieces of research, such as a study by Ebert and Steinert [47], which examined experiences of IPV during COVID-19 in Germany. Responses were dichotomously recoded, where 1 indicated any experience of each subsequent type of IPV and 0 indicated no experience of each type of IPV in the past month [48].	5
Men's perpetration of IPV	Items adapted from the WHO VAWI and the core questionnaire in the WHO-MCS [45]: physical (1 item), psychological (1 item), sexual (1 item), coercion (1 item), and economic abuse (1 item). Participants were asked to give a frequency score on a scale of 0 to >8 times, with responses dichotomized into any perpetration of IPV in the past month.	5
Secondary behavioral outcomes		
Men's experiences of IPV	Two items used and adapted from Abramsky et al [49], namely: "In the past month, during any potential times that you may have used violence against your partner, did they ever fight back physically to defend themselves?" and "In the past month, have you ever been hit or physically mistreated by your partner when you were not hitting or physically mistreating them?" Respondents were asked to give a frequency score on a scale of 0 to 8 or more times, with responses dichotomized.	2
Women's perpetration of IPV	Two items were adapted from a questionnaire developed by researchers at the London School of Hygiene and Tropical Medicine for a randomized controlled trial of a violence prevention intervention in Tanzania [49]. Items included: "During any potential times that you were hit in the past month, did you ever fight back physically to defend yourself?" and "In the past month, have you ever hit or physically mistreated your partner when they were not hitting or physically mistreating you?" Respondents were asked to give a frequency score on a scale of 0 to 8 or more times, with responses dichotomized.	2
Women's positive partner interactions	Item adapted from Abramsky et al [49] asking: "How many times in the past month did your partner show you they cared and respected your feelings even though they disagreed with you?" Participants were asked to give a frequency score on a scale of 0 to 8 or more times.	1
Men's positive partner interactions	One item adapted from Abramsky et al [49], that asks, "How many times in the past month did you show your partner you cared and respected their feelings even though you disagreed?." Participants asked to give a frequency score on a scale of 0 to 8 or more times.	1

Outcomes	Study measurement	Items, n
Gender-equitable behaviors	Measured using items adapted from questionnaires used in previous violence prevention interventions [23,49]. In total, 3 items were adapted from Abramsky et al [49], which included questions on couple communication, joint decision-making, and partner conflict resolution, with questions such as: "In the past week, how many times did you and your partner talk about your worries and feelings?" and "In the past week, how many times did you and your partner make a decision together?" The fourth item was adapted from a questionnaire developed by researchers of a randomized controlled trial of a gender-transformative violence prevention intervention in Rwanda [23], which asks: "In the past week, how many times did you and your partner share housework and caregiving tasks equally?" Respondents were asked to report whether in the past week the specific behavior occurred on a scale of 0 (never) to 3 (many times). The scores were summed for all items for the overall gender-equitable behavior score. The possible overall score ranged from 0 to 12.	4
Attitudes toward gender roles and IPV	Measured using 6 items in the "Attitudes toward gender roles" section of the WHO-MCS [45], such as "A woman should obey her husband's wishes even if she disagrees." Respondents were asked to indicate whether they agree or disagree with the statements using a Likert scale ranging from 1 (strongly agree) to 5 (strongly disagree). Negative items were reverse-coded so that higher scores denoted more gender-equitable attitudes. The scores were summed for all items for the overall attitude score. The possible overall score ranged from 6 to 30.	6
Background variables		
Parent gender, age, relationship status, living with partner, education level, employment status, partner employment status, and number of children	Sociodemographic information	8 (1 item each)
Women's past-year IPV experience (measured at baseline)	Measured using 1 item adapted from the WHO VAWI and the core questionnaire in the WHO-MCS [45], which asks, "How many times, in the past 12 months, did your partner do one of the following: insult or shout at you, hit or shove you, or force you to have sex?" Respondents were asked to give a frequency score on a scale of 0 to 8 or more times. Due to the skewed response patterns in the data for IPV, responses for IPV were dichotomously recoded, where 1 indicated any experience of IPV and 0 indicated no experience of IPV in the past year [48].	1

[a]IPV: intimate partner violence.

[b]WHO: World Health Organization.

[c]VAWI: Violence Against Women Instrument.

[d]MCS: Multi-Country Study on Domestic Violence.

[e]ISPCAN: International Society for Prevention of Child Abuse and Neglect.

Statistical Analysis

Data analyses were conducted using R (R Core Team). Descriptive statistics and frequencies were used to describe the sociodemographic characteristics of participants and to report engagement outcomes, summarized using means (SDs), medians (range), and N (%). To investigate changes in outcomes between pre and posttest, McNemar chi-square tests (for dichotomous variables) and paired dependent-sample *t* tests (for continuous variables) were conducted, with corresponding 95% CIs and *P* values [50,51]. Assumptions were examined to confirm no violations existed. Data were analyzed using complete case analysis.

Ethical Considerations

Ethical procedures were followed to ensure the safety and welfare of study participants. At the start of the IPV assessment, screening procedures ensured that men and women from the same household were not interviewed about IPV. This is based on the World Health Organization VAW research guidelines [52] and ethical and safety recommendations for intervention research on VAW [45] to prevent putting women at risk. Consequently, men who responded that their partner was taking part in the intervention were not asked the IPV items in the IPV assessment and were only asked instead the questions on attitudes toward IPV and gender roles and on gender-equitable behaviors. In addition, efforts were also in place to ensure responses to the IPV questions could not be viewed by others by delivering the IPV assessment using ODK. This procedure was implemented to protect participants by ensuring no responses were saved on their mobile devices. Participants who disclosed experiences of IPV were automatically provided referrals to local services, customized to their regional context that supported individuals experiencing violence. Local referral information could also be accessed by writing "Help Me" and choosing "Other Support" in the ParentText free text field. In line with UNICEF risk communication and community engagement guidelines [53], ParentText is also formatted to detect high-risk keywords to identify potential disclosure of dangerous situations via the free text field. Following detection, ParentText automatically offers the participant an empathetic and empowering reply with referral details that are customized to the country, which supports parent and child safety (eg, hotlines, police, and ambulance).

Further ethical considerations that were taken within the research project include recognitions surrounding power and limitations of positionality [54]. For instance, it is worth noting that while collaborators and members of the research team came from several countries, the primary author of this paper is from and is part of a research institution in the global north. Efforts to address limitations associated with this positionality were put in place, including integrating reflexivity across the stages of the research process [55]. For example, during the research design and data collection planning phase, members of local grass-roots organizations, communities, and stakeholders in Jamaica and South Africa were continuously consulted to review and revise the IPV prevention content and assessment materials [28]. Integrating these embedded consultations and opportunities for research amendments with members of the local community was a pivotal part of the research project, particularly given the first author's positionality as an "outsider" [55].

Ethics approval was obtained by the Department Research Ethics Committee at the University of Oxford (Central University Research Ethics Committee 2 Ref No: R69569/RE009). All methods were carried out in accordance with relevant guidelines and regulations. Informed consent to participate in the study was obtained from all participants before enrolling in the study and the intervention.

Results

Participants Characteristics

Participant characteristics are provided in Table 2 by country and with an overall summary. A study flow diagram illustrating the flow of participants through the trial is provided in Figure 5. The average age of participants across both countries was 34.4 (SD 8.74; range 16-57) years. The average age of women in South Africa was 34.9 (SD 6.77; range 25-49) years and men 28.8 (SD 2.68; range 25-36) years. In Jamaica, where only women were recruited into the study, the average age of women was 39.9 (SD 11.52; range 16-57) years. Most participants, across both countries, were partnered but not married (48/68, 71%). On average, participants had 2.23 (SD 1.19) children. Approximately half of the participants across both countries had a secondary education (36/68, 53%), and 7% (5/68) had only primary education or less. In South Africa, among women participants, 42% (8/19) were working and 58% (11/19) were unemployed or looking for work. Among men, slightly more were working (13/21, 62%), with 29% (6/21) unemployed or looking for work. In Jamaica, among women participants, this was similar, with 43% (12/28) working and 54% (15/28) unemployed or looking for work. Across both countries, the employment of partners was high (50/68, 74%), with only 18% (12/68) with partners who were unemployed or looking for work. Experience of IPV (women-report only due to safety reasons) at baseline was high: 21% (4/19) of women participants in South Africa and 29% (8/28) of women in Jamaica had experienced some form of IPV in the past year, respectively.

Table 2. Demographic characteristics of study participants in South Africa and Jamaica.

	South Africa		Jamaica	Overall
	Women (n=19)	Men (n=21)	Women (n=28)	Total (N=68)
Gender, n (%)				
Women	19 (48)	N/A[a]	28 (100)	47 (69)
Men	N/A	21 (53)	—[b]	21 (31)
Age (y), mean (SD)	34.9 (6.77)	28.8 (2.68)	39.9 (11.52)	34.4 (8.74)
Relationship, n (%)				
Married	6 (32)	2 (10)	12 (43)	20 (29)
Partnered but not married	13 (68)	19 (90)	16 (57)	48 (71)
Number of children, mean (SD)	2.53 (1.17)	1.19 (0.68)	2.61 (1.25)	2.23 (1.19)
Education, n (%)				
Primary	1 (5)	0 (0)	4 (14)	5 (7)
Secondary	9 (47)	10 (48)	17 (61)	36 (53)
Higher	9 (47)	11 (52)	7 (25)	27 (40)
Employment, n (%)				
Working	8 (42)	13 (62)	12 (43)	33 (49)
Unemployed or looking for work	11 (58)	6 (29)	15 (54)	32 (47)
Student	0 (0)	2 (10)	0 (0)	2 (3)
Retired	0 (0)	0 (0)	1 (4)	1 (2)
Partner's employment, n (%)				
Working	11 (58)	17 (81)	22 (79)	50 (74)
Unemployed or looking for work	7 (37)	2 (10)	3 (11)	12 (18)
Student	0 (0)	2 (10)	0 (0)	2 (3)
Retired	0 (0)	0 (0)	2 (7)	2 (3)
Refuse to answer	1 (5)	0 (0)	1 (4)	2 (3)
Experience of IPV[c], n (%)				
Experienced IPV in the past year[d]	4 (21)	X[e]	8 (29)	12 (26)[f]

[a]N/A: not applicable.

[b]No men were recruited in Jamaica.

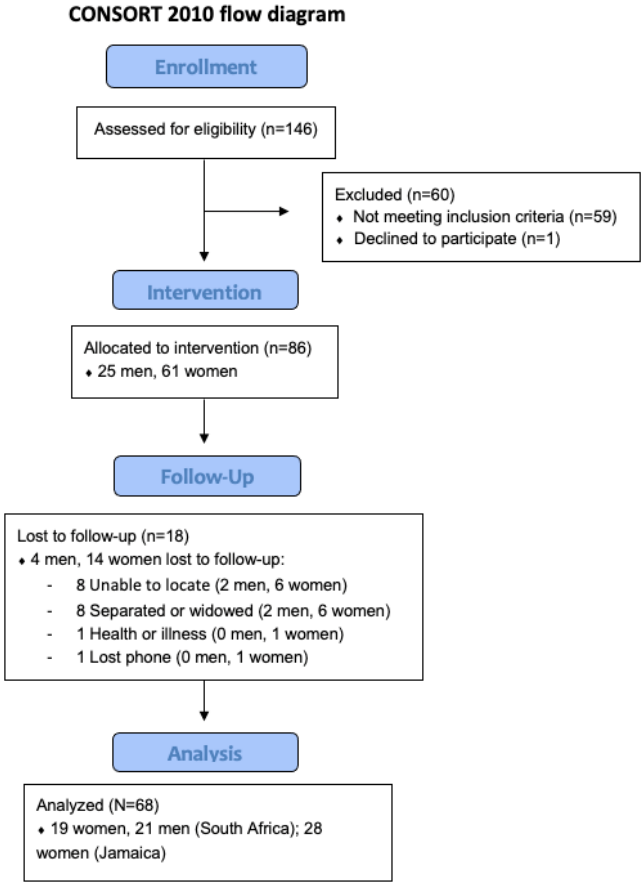
[c]IPV: intimate partner violence.

[d]For safety reasons, men were not asked this question at baseline.

[e]Men were not asked this question at baseline due to safety reasons.

[f]Overall IPV here only includes women.

Figure 5. Study flow diagram illustrating the flow of participants through the trial. CONSORT: Consolidated Standards of Reporting Trials.



User Engagement Outcomes

Overall engagement outcomes are provided in Table 3. The ParentText program lasted 23 days, with the main IPV prevention content delivered on day 2 and check-in messages delivered within the first 12 days (Multimedia Appendix 2 provides an overview of the program content delivered each day). In terms of engagement, in South Africa, the average number of days in the program was 13 (SD 23.90) and 3 (SD 3.78) days for women and men, respectively. The average number of days users interacted with the chatbot was 3 (SD 1.58) for women and 2.6 (SD 1.52) for men, each with an average interaction rate of 0.57 (SD 0.26) interactions and 0.59 (SD 0.17) interactions per day, respectively. In Jamaica, the average number of days in the program was 27 (SD 23.10) days, and the average total number of days users interacted with the chatbot was 5.2 (SD 5.45) days. The average interaction rate in Jamaica was 0.21 (SD 0.20) interactions per day. In terms of engagement across the 5 relationship topics, there was a greater engagement with topics in Jamaica where engagement with each topic ranged from 11% (3/28) to 29% (8/28) compared to South Africa, where it ranged from 0% to 11% (2/19). In terms of completion rates, in South Africa, 25% (5/20) and 5% (1/20) of women and men completed 1 out of the 5 IPV topics, and 5% (1/20) and 0% completed all 5 IPV topics, respectively. In comparison to the completion rate of overall ParentText modules, 16% (3/19) and 9% (2/21) of women and men completed all the parenting (ie, the non-IPV) ParentText modules in South Africa. In contrast, in Jamaica, 21% (6/28) completed 1 out of the 5 IPV topics, 11% (3/28) completed all 5 IPV topics, and 18% (5/28) completed all the parenting ParentText modules.

Table 3. ParentText user engagement outcomes of the 23-day program in South Africa and Jamaica, including overall retention, engagement per intimate partner violence (IPV) topic, and IPV topic completion rate.

Outcomes	South Africa		Jamaica
	Women (N=19)	Men (N=21)	Women (N=28)
Overall retention, mean (SD)			
Number of days in program[a]	13 (23.90)	3 (3.78)	27 (23.10)
Number of active days the user interacted with the chatbot[b]	3 (1.58)	2.6 (1.52)	5.2 (5.45)
Number of interactions per day[c]	0.57 (0.26)	0.59 (0.17)	0.21 (0.20)
Engagement per IPV topic[d]			
Participants who viewed each IPV topic, n (%)			
Relationship topic 1: “Treat each other as equals”	2 (11)	2 (10)	7 (25)
Relationship topic 2: “Become a confident parent and supportive spouse”	2 (11)	0 (0)	4 (14)
Relationship topic 3: “Share family responsibilities”	2 (11)	1 (5)	5 (18)
Relationship topic 4: “Resolve conflict peacefully”	2 (11)	1 (5)	3 (11)
Relationship topic 5: “Listen and talk to each other”	2 (11)	1 (5)	8 (29)
All relationship topics (viewed all relationship topics)	1 (5)	0 (0)	3 (11)
Completion rate of IPV topics completed, n (%)[e]			
1	5 (25)	1 (5)	6 (21)
2	0 (0)	0 (0)	3 (11)
3	0 (0)	0 (0)	1 (4)
4	0 (0)	1 (5)	0 (0)
5	1 (5)	0 (0)	3 (11)
Completion rate of ParentText parenting modules, mean (SD)	16 (0.19)	9 (0.06)	18 (0.20)

[a]Measured by the final day participant used the chatbot, subtracted by the first day the participant used the chatbot (note: this measure includes inactive days).

[b]Number of days participants actively interacted with ParentText (note: this measure excludes inactive days).

[c]Measured by the number of interactions logged per day divided by the total number of days the participant actively interacted with ParentText.

[d]Measured based on the number of participants who viewed each IPV topic.

[e]Number of IPV topics completed out of the total 5 topics.

Behavioral Outcomes

Experiences and Perpetration of IPV

Table 4 presents the descriptive results of the IPV outcomes, summarizing experiences of IPV among women and perpetration of IPV among men using past-month frequencies at pre and posttest and *P* values for each outcome. At posttest, 73% (11/15) of women in South Africa and 72% (13/18) of women in Jamaica reported experiencing some form of IPV, compared to 84% (16/19) and 96% (23/24) at baseline, reductions that were significant in South Africa (P=.01) and in Jamaica (P=.01). Experiences of all subtypes of IPV for women in both South Africa and Jamaica also showed reductions at posttest, and so did acts of self-defense and perpetration (Multimedia Appendix 5). For perpetration of IPV among men, at posttest, 75% (9/12) of men in South Africa perpetrated some form of IPV, compared to 85% (11/13) at baseline, which was nonsignificant for overall perpetration of IPV (P=.06). Analyses of subtypes of IPV, however, found significant reductions in psychological (P=.04), physical (P<.001), and sexual (P<.001) violence among men in South Africa (Multimedia Appendix 5). Detailed tables of the primary and secondary outcomes are provided in Multimedia Appendix 5.

Table 4. Primary behavioral outcomes for women's past-month experience of overall intimate partner violence (IPV) in South Africa and Jamaica and men's past-month perpetration of overall IPV in South Africa.

	Baseline overall IPV, n (%)[a]	Posttest overall IPV, n (%)	*P* value[b]
Women's experience of overall IPV (past month)			
South Africa	16/19 (84)	11/15 (73)	.01
Jamaica	23/24 (96)	13/18 (72)	.01
Men's perpetration of overall IPV (past month)			
South Africa	11/13 (85)	9/12 (75)	.06

[a]% used is proportional to the baseline and posttest samples, respectively.

[b]McNemar chi-squared *P* value.

Positive Partner Interactions

Table 5 presents outcomes on positive partner interactions measured through respectful behaviors, summarized using means and SDs for pre and posttest along with *P* values. There were no significant effects detected for change in positive partner interactions reported by men or women.

Table 5. Secondary behavioral outcomes for positive partner interactions in the past month, overall attitude toward gender roles and intimate partner violence (IPV), and overall gender-equitable behaviors in the past week for women and men in South Africa and women in Jamaica.

	Baseline, mean (SD)	Posttest, mean (SD)	*P* value (*t* test)
Positive partner interactions			
South Africa			
Women (n=15)	4.74 (2.84)	4.14 (2.98)	.57
Men (n=12)	3.92 (2.50)	5.67 (2.46)	.10
Jamaica			
Women (n=15)	5.08 (2.60)	3.89 (2.54)	.14
Overall attitude[a] toward gender roles and IPV			
South Africa			
Women (n=14)	10.11 (2.71)	10.13 (3.89)	.98
Men (n=16)	11.90 (3.82)	8.88 (2.63)	.01
Jamaica			
Women (n=15)	11.39 (3.45)	11.89 (2.35)	.56
Overall gender-equitable behaviors[b]			
South Africa			
Women (n=15)	7.37 (2.95)	7.93 (3.53)	.63
Men (n=16)	7.81 (1.94)	9.75 (1.61)	.02
Jamaica			
Women (n=15)	7.14 (3.49)	7.39 (2.85)	.80

[a]A lower score indicates less harmful attitudes toward IPV and gender roles.

[b]A higher score indicates more gender-equitable behaviors in the past week.

Attitudes Toward (Harmful) Gender Roles and IPV

Table 5 summarizes means and SDs for the attitude outcomes at pre and posttest, with *P* values for each attitude outcome also provided. Analyses indicate a significant reduction in harmful attitudes toward IPV and gender roles reported by men in South Africa (*P*=.01); however, no effects were reported among women in South Africa (*P*=.98) or in Jamaica (*P*=.56).

Gender-Equitable Behaviors

Analyses found a significant increase in gender-equitable behaviors for men (*P*=.02; Table 5). No significant effects were detected for women in South Africa (*P*=.63) or in Jamaica (*P*=.80).

Discussion

Principal Findings

To the best of our knowledge, this pre-post pilot study is the first that examines user engagement and indicative outcomes in a digital parenting intervention with integrated IPV prevention content [22]. The study offers valuable, preliminary insights across 2 countries on user engagement with a digital parenting intervention with integrated IPV prevention, as well as provides exploratory, indicative results on IPV experience and perpetration, attitudes toward IPV, and gender-equitable behavior outcomes. First, we will discuss the key findings on levels of user engagement. While the average daily interaction rate was higher in South Africa (0.57 and 0.59 interactions per day for women and men, respectively) compared to Jamaica (0.21 interactions per day), the average number of days users stayed in the program in Jamaica was greater (27 days), compared to South Africa (13 and 3 days for women and men, respectively). Due to WhatsApp messaging restrictions, "push" messaging was paused after 24 hours of user inactivity, until user activity resumed. Given research findings that highlight the importance of messaging reminders in digital interventions [56], this restriction may have contributed to some of the low engagement levels observed.

The completion rate of engaging with at least 1 IPV topic in the program was lower than expected across the intervention, with completion rates of 5% (1/20) for men and 25% (5/20) for women in South Africa, and 21% (6/28) for women in Jamaica. The completion rates of all 5 IPV topics were much lower, with 0% of men and 5% (1/20) of women in South Africa, and 11% (3/28) of women in Jamaica completing all 5 IPV prevention topics. Even though user engagement and retention rates have been noted as a challenge in digital interventions [57], completion rates in this study were lower than anticipated in comparison to findings in the field. For example, a review of digital interventions for parents with young children found that attrition rates in digital parenting programs ranged from 30% to 50% [58]. Accordingly, the low completion rates of the IPV topics observed in ParentText raise questions about whether user engagement in the program was sufficient to change complex behaviors and deeply rooted attitudes surrounding violence and gender roles.

There are various possible reasons for the low levels of engagement with the IPV content observed in ParentText. First is the format of the program schedule. On day 2 of the ParentText program, participants were provided with a list of all 5 IPV prevention topics, which they could choose to explore. After selecting a topic to view, users were given the option to view one of the other IPV topics. However, it is possible that viewing >1 IPV prevention topic per day may have been overwhelming for users. Therefore, this may have inhibited users from engaging with multiple IPV topics in the program. Another element that may have affected engagement was the timing of the program "check-in" reminders. Participants received "check-in" messages in the latter part of the program (days 4-12; Multimedia Appendix 2), where they were encouraged to revisit and explore the IPV topics they had not yet engaged with. However, given the already low average number of days for which the participants remained in the program, it is likely that many participants did not receive these messages. These findings raise important questions on how to increase engagement with the IPV modules integrated into the ParentText. For example, amending the timing of content delivery might be one way to increase engagement. For example, a study assessing a text-messaging intervention focusing on promoting health relationships found that the timing of message delivery played a key role in program engagement and that if delivered at the wrong time, users might be less likely to engage with the text messages [59].

Taken together, these study results give valuable insight into how users engage with IPV content integrated into a chatbot and suggest that rather than interacting with the chatbot content daily, users appear to interact with the chatbot at their own pace. Consequently, providing more options for user-led selection of program content would be important for future iterations of the chatbot to ensure users can access and revisit the information they need more readily. The importance of user-led content delivery has also been underscored by other digital parenting intervention research. For instance, findings from a study on the digital fatherhood program "SMS4dads" revealed that the length of time between users receiving a text message with a link to program content to that of users clicking the link was, on average, 2.11 (SD 3.94) days [60]. This highlights how engagement with digital intervention content is often not instantaneous, but rather asynchronous and driven by users' schedules [60]. Accordingly, in future versions of ParentText, combining "push" notification reminders with user-led selection of program content might help increase user engagement. Notably, since the present trial of ParentText in South Africa and Jamaica, newer versions of the WhatsApp settings have been released, which allow future iterations of the program to override the 24-hour limit inhibiting "push" messages after user inactivity.

More broadly, the relatively low levels of user engagement found in this pilot study raise questions as to whether a digital modality is sufficient when delivering intervention content related to complex gender-equitable beliefs and behaviors. It may be the case that a hybrid approach, combining both digital content delivery and in-person sessions, might be necessary. For instance, a study on enhancing father engagement in parenting programs revealed that in terms of intervention delivery, fathers prefer less intensive and low-dose programs, including internet - based interventions [61]. However, a recent systematic review of interventions promoting gender equality noted that in various studies it was found that dialog was critical for driving change in gender norms [62]. The importance of dialog is further underscored by research examining men's engagement in family and health interventions, revealing that men value and often seek the opportunity to connect with other men to discuss parenting and fatherhood, even though cultural norms sometimes hinder this. Together, these findings suggest that a combination of digitally delivered content and in-person discussion opportunities might be needed. Various studies also suggest that using face-to-face recruitment and participatory, peer-engaged methods is an effective approach for engaging

parents, particularly fathers [63,64]. Research findings also suggest that providing opportunities to critically discuss and engage in dialog with others, especially on difficult topics, plays a valuable role not only in terms of engagement but also in terms of creating change and shifting harmful gender attitudes [17,62,64]. Taken together, it may be possible that low-intensity, 1-way engagement with a chatbot is insufficient to create change in more deeply rooted behaviors and attitudes among users and that more interactive engagement is needed to shift behaviors and beliefs. While these hybrid options and amendments may improve engagement, it should be noted that they will also increase intervention costs and limit scalability, which are some of the benefits of digital-only interventions [24].

It is also possible that the low level of user engagement with the IPV prevention topic was due to other factors, such as the format of the program. The IPV content in ParentText was only delivered on 1 specific day, after which only check-in messages were sent (refer to Figure 4 for an overview of the content delivery schedule). Receiving the option to view all the IPV content on the same day may potentially have been overwhelming for participants. This may have reduced user engagement rates due to intervention fatigue [65], which is cognitive or emotional weariness resulting from competing demands of intervention engagement and other burdens in daily life [66] and has been linked to intervention adherence. Considerations for future iterations of ParentText include (1) delivering all the content automatically (since optional and supplemental content in digital interventions is often associated with low engagement [67]) and (2) adjusting the timing of content and spreading out the delivery of the IPV content to allow mental rests in between new material, strategies that have been suggested to reduce intervention fatigue [65]. Notably, this is already an amendment that future iterations of ParentText are incorporating and that is currently being implemented in a trial of an updated version of ParentText in Malaysia [68]. This integrated approach of implementing IPV prevention content beyond a single module is also a strategy that is favored and advocated by gender-transformative approaches and is one that is being advocated more broadly in the parenting field [69].

The exploratory analyses of the behavioral outcomes, while only preliminary, are also important to discuss. In terms of IPV, women in Jamaica reported a significant reduction in overall experiences of IPV at posttest. These tentative findings are in line with results from other parenting and gender-transformative programs that have found reductions in both IPV experience at posttest and even at follow-up [18,23,37]. For example, findings from an RCT of a gender-transformative intervention in Rwanda found that, compared to the control, women in the intervention group reported reductions in both physical and sexual IPV following the intervention [23].

In contrast to the significant reductions in IPV experience among women, no significant effects for overall IPV perpetration were detected among men in South Africa. However, analyses of IPV subtypes found significant reductions in psychological violence, physical violence, and sexual violence among men in South Africa (Multimedia Appendix 5). The lack of change in overall IPV perpetration might be related to the low level of male engagement. Notably, the user engagement and the IPV completion rate of men in South Africa were 50% less than those of women. Hence, the limited interaction men had with the chatbot may not have been sufficient to change these behaviors.

It may also be the case that more interactive engagement is needed to shift these complex behaviors. For instance, emerging findings suggest that content modality can impact digital intervention outcomes [42], as highlighted in a recent study on ChattyCuz, a chatbot intervention that aims to support young women in navigating intimate relationships in South Africa [70]. Interestingly, an RCT of ChattyCuz found that while the gamified version of the chatbot (treatment 1) led to modest but significant reductions in IPV experience compared to those without treatment, in the narrative treatment, there was no effect on IPV experience [70]. The authors hypothesize that the lack of measurable effect on IPV experience in the narrative treatment arm of the study might be related to the lack of user interaction, that greater engagement with intervention content is necessary for behavior change, and that an information-only approach is insufficient [70]. These findings highlight how low levels of user engagement in ParentText might play a role in the lack of behavior changes detected and underscore that, amendments to the program modality might be needed to make the content more interactive to shift behaviors. Recent studies have also demonstrated the strong impact that interpersonal relationships have on gender attitudes, underscoring the need for interventions to not only target individuals but also social networks, such as peers and family members through, for example, critical reflection and dialog among social groups [71]. Research suggests that programs that have shown evidence of change in gender norms involve the engagement of stakeholders at different levels of the social-ecological model, include activities that allow for active participation, and promote critical awareness [14]. Consequently, it is possible that greater interactive engagement is needed to see greater changes in IPV experience and perpetration.

This study also revealed a significant overall reduction in harmful attitudes toward IPV as well as a significant increase in overall gender-equitable behaviors among men in South Africa. These indicative effects are promising, and while caution is warranted given the small sample, the findings align with other interventions that have observed reductions in harmful attitudes among men at posttest and detected increases in gender-equitable behaviors following gender-transformative interventions [72,73].

Strengths and Limitations

The study has various strengths worth noting. As the first study of a digital parenting intervention with integrated IPV prevention content, the present research makes a valuable contribution to the field of violence prevention and targeted efforts seeking to address VAC and IPV concurrently. With a growing emphasis in policy and research on preventing multiple forms of violence simultaneously [74], this study sheds light on preliminary behavioral outcomes and how users engage in IPV prevention content that has been integrated into a digital parenting intervention. Another strength of the study is the use of data from 2 countries. By delivering the program in 2 different

regions and cultures, this study offers an insight into important considerations to bear in mind before scaling up the intervention. For instance, the challenges with recruiting men in Jamacia compared to South Africa suggest that alternative recruitment strategies might be needed when scaling up the intervention in the Caribbean region.

The study also has various limitations. The first relates to the lack of a control group, without which causality cannot be established from the pre-post data. The study also had a small sample size, which means that the research findings need to be interpreted with caution given that there is limited statistical power in detecting pre-post intervention effects. Consequently, the next step is a fully powered RCT to establish effectiveness. Future research would also benefit from incorporating a factorial experiment design, which can be used to examine which specific intervention components have an effect on outcomes and which components interact with each other [75]. Knowing which components should be kept to sustain impact when taking an intervention to scale or delivering it in a new setting is critical, especially when considering program length and implementation costs [76]. Including follow-up assessments to examine whether intervention effects remain, become reduced, or are delayed over time [77] would also be beneficial in future studies as this study was limited to a 6-week posttest due to resource and time constraints. In addition, the study used a convenience sampling approach, which may limit the generalizability of the study findings. For example, it is possible that caregivers who opted to take part in the parenting program were more willing to change, which may bias the results. The outcomes in the study were also self-reported, and consequently, they may be subject to disclosure bias, recall bias, and social-desirability bias, leading to under reporting of unfavorable behaviors and over reporting of what might be perceived as desirable answers. However, given that the self-reports were carried out on participants' phones and participants were told the reports were fully anonymized, this will likely have reduced desirability bias in the quantitative surveys.

Another important limitation is that while in this study, participants were provided with mobile credit, this may not always be realistic in a real-world setting. While access to the internet and mobile data are increasing globally [78], it is important to explore potential options of being able to use the program without mobile data, especially when planning strategies for scale-up and delivery. One digital intervention that has explored this is the parenting program ParentApp for Teens, which is a mobile app designed for offline use that targets parents of teenagers [79,80]. There are also assessment limitations related to the self-reported measurements used in this study. Given the sensitive nature surrounding IPV and attitudes toward violence, it is possible that participants were hesitant to disclose violence due to stigma or fear of being endangered, leading to under reporting and consequently leading to undetected changes in behaviors [81]. Similarly, social-desirability bias may have also impacted the measurement of behavioral outcomes, particularly the measurement of IPV perpetration, which has been found to be more prone to social-desirability bias than reports of IPV experience [82].

An additional limitation worth noting is that men were underrepresented in the study. Recruitment of male caregivers is a known difficulty with parenting interventions [83], and even though it was possible to recruit both male and female caregivers in South Africa, it was only possible to recruit female caregivers in Jamaica. The lack of male caregivers from Jamaica means that certain experiences may have been underestimated due to the low number of fathers included. This challenge highlights the need for future studies to ensure greater efforts are taken to adopt recruitment strategies that ensure men are also included. In future studies, it would also be valuable to include people aged between 16 and 17 years in the South African sample to better understand how adolescent parents in South Africa engage with the chatbot.

Conclusions

This research presents tentative findings on the first-ever attempt to research the integration of IPV and parenting content into a digital intervention. Some of the preliminary results of the intervention are promising, with indications of reductions in overall IPV for women and a potential decrease in harmful IPV attitudes among men. However, the high rates of IPV experiences and perpetration reported at the posttest are still concerning and suggest that greater intervention efforts and programmatic amendments are likely necessary to tackle deep-rooted attitudes and harmful behaviors. In addition, the unexpectedly low levels of engagement in the intervention raise questions about whether interactions with the chatbot were sufficient to shift behaviors and attitudes and suggest that a hybrid approach to program delivery might be necessary. Given the urgent need for scalable prevention efforts that can tackle multiple forms of violence concurrently [84], more research is necessary to further explore how to increase chatbot user engagement and achieve greater reductions in IPV and harmful attitudes toward gender roles and IPV.

Acknowledgments

This research was funded in whole, or in part, by the United Kingdom Research and Innovation (UKRI; grant ES/P000649/1). For the purpose of open access, the author has applied for a CC BY public copyright license to any author accepted manuscript version arising from this submission. This research was also funded by UNICEF (United Nations Children's Fund) Jamaica and the Economic and Social Research Council Grand Union Doctoral Training Partnership in collaboration with the University of Oxford Saven European Scholarship and the Balliol College Dervorguilla Scholarship in Social Sciences. This study is part of the Global Parenting Initiative, which is funded by the LEGO Foundation, Oak Foundation, the World Childhood Foundation (grant 16191), the Human Safety Net, and the UK Research and Innovation Global Challenges Research Fund (grant ES/S008101/1). Aside from UNICEF's involvement in guiding the selection of the research site and study time frame, none of the funders were

involved in study design, data collection, data analysis and interpretation, or the writing of the manuscript. The authors would like to thank those who made this research possible: the respondents who participated in this study, UNICEF Jamaica, the research assistants, the study funders, the researchers and stakeholders who provided input during intervention development and delivery, and the Department of Social Policy and Intervention at the University of Oxford for administrative support. Generative artificial intelligence was not used in any portion of the manuscript writing.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

MS, JL, and FG participated in the design of the study and oversaw data collection and cleaning. MS managed the data collection process, with support from PZ, FACA, GHR, and RS. MS designed the statistical analysis, with JL, FG, QH, FACA, CF, and LC providing guidance on the analysis approach. MS drafted the paper, with JL and FG contributing. All authors approved the final manuscript.

Conflicts of Interest

MS is a researcher at the University of Oxford, and JL and FG are employed as staff at the university. In addition, JL is the chief executive officer of Parenting for Lifelong Health, a charitable organization based in the United Kingdom that developed the ParentText intervention. MS, JL, PZ, QH, FACA, GHR, and FG have participated and are participating in several research studies involving the program as investigators at the universities of Oxford, Cape Town, and Glasgow. CF and LC are employees of IDEMS International, a technology development company that codeveloped the ParentText intervention and provided support on data analysis of user engagement. RS is an employee at UNICEF and provided support on data collection. Conflict is avoided by declaring these potential conflicts of interest and by conducting and disseminating rigorous, transparent, and impartial evaluation research on both this and other similar parenting programs. No profits will be made from this program.

Multimedia Appendix 1
Intimate partner violence assessment.
[DOCX File , 28 KB - formative_v9i1e58611_app1.docx]

Multimedia Appendix 2
Structure of the ParentText intervention.
[DOCX File , 24 KB - formative_v9i1e58611_app2.docx]

Multimedia Appendix 3
Overview of ParentText intimate partner violence prevention content.
[DOCX File , 20 KB - formative_v9i1e58611_app3.docx]

Multimedia Appendix 4
CONSORT (Consolidated Standards of Reporting Trials) checklist.
[DOCX File , 374 KB - formative_v9i1e58611_app4.docx]

Multimedia Appendix 5
Detailed primary and secondary outcome tables.
[DOCX File , 46 KB - formative_v9i1e58611_app5.docx]

References

1. Violence against women prevalence estimates, 2018. World Health Organization. 2021 Mar 9. URL: https://www.who.int/publications/i/item/9789240022256 [accessed 2024-12-10]
2. Whitten T, Tzoumakis S, Green MJ, Dean K. Global prevalence of childhood exposure to physical violence within domestic and family relationships in the general population: a systematic review and proportional meta-analysis. Trauma Violence Abuse 2024 Apr 10;25(2):1411-1430 [FREE Full text] [doi: 10.1177/15248380231179133] [Medline: 37300288]
3. Hillis S, Mercy J, Amobi A, Kress H. Global prevalence of past-year violence against children: a systematic review and minimum estimates. Pediatrics 2016 Mar;137(3):e20154079 [FREE Full text] [doi: 10.1542/peds.2015-4079] [Medline: 26810785]
4. Global status report on preventing violence against children 2020. World Health Organization. URL: https://www.unicef.org/media/70731/file/Global-status-report-on-preventing-violence-against-children-2020.pdf [accessed 2024-12-10]

5. Global and regional estimates of violence against women: prevalence and health effects of intimate partner violence and non-partner sexual violence. World Health Organization. 2013 Oct 20. URL: https://www.who.int/publications/i/item/9789241564625 [accessed 2024-12-10]
6. Cerna-Turoff I, Fang Z, Meierkord A, Wu Z, Yanguela J, Bangirana CA, et al. Factors associated with violence against children in low- and middle-income countries: a systematic review and meta-regression of nationally representative data. Trauma Violence Abuse 2021 Apr;22(2):219-232 [FREE Full text] [doi: 10.1177/1524838020985532] [Medline: 33461441]
7. Pearson I, Page S, Zimmerman C, Meinck F, Gennari F, Guedes A, et al. The co-occurrence of intimate partner violence and violence against children: a systematic review on associated factors in low- and middle-income countries. Trauma Violence Abuse 2023 Oct;24(4):2097-2114 [FREE Full text] [doi: 10.1177/15248380221082943] [Medline: 35481390]
8. Guedes A, Bott S, Garcia-Moreno C, Colombini M. Bridging the gaps: a global review of intersections of violence against women and violence against children. Glob Health Action 2016 Jun 20;9:31516 [FREE Full text] [doi: 10.3402/gha.v9.31516] [Medline: 27329936]
9. Alhusen JL, Ho GW, Smith KF, Campbell JC. Addressing intimate partner violence and child maltreatment: challenges and opportunities. In: Korbin J, Krugman R, editors. Handbook of Child Maltreatment. Dordrecht, The Netherlands: Springer; 2014.
10. Gracia E, Rodriguez CM, Martín-Fernández M, Lila M. Acceptability of family violence: underlying ties between intimate partner violence and child abuse. J Interpers Violence 2020 Sep;35(17-18):3217-3236. [doi: 10.1177/0886260517707310] [Medline: 29294751]
11. Chisholm CA, Bullock L, Ferguson JE. Intimate partner violence and pregnancy: epidemiology and impact. Am J Obstet Gynecol 2017 Aug;217(2):141-144. [doi: 10.1016/j.ajog.2017.05.042] [Medline: 28551446]
12. Guedes A, Mikton C. Examining the intersections between child maltreatment and intimate partner violence. West J Emerg Med 2013 Aug;14(4):377-379 [FREE Full text] [doi: 10.5811/westjem.2013.2.16249] [Medline: 23997846]
13. Edwards KM, Kumar M, Waterman EA, Mullet N, Madeghe B, Musindo O. Programs to prevent violence against children in sub-Saharan Africa: a systematic review. Trauma Violence Abuse 2024 Jan;25(1):593-612 [FREE Full text] [doi: 10.1177/15248380231160742] [Medline: 36964686]
14. Levy JK, Darmstadt GL, Ashby C, Quandt M, Halsey E, Nagar A, et al. Characteristics of successful programmes targeting gender inequality and restrictive gender norms for the health and wellbeing of children, adolescents, and young adults: a systematic review. Lancet Glob Health 2020 Feb;8(2):e225-e236 [FREE Full text] [doi: 10.1016/S2214-109X(19)30495-4] [Medline: 31879212]
15. Brush LD, Miller E. Trouble in paradigm: "gender transformative" programming in violence prevention. Violence Against Women 2019 Nov;25(14):1635-1656. [doi: 10.1177/1077801219872551] [Medline: 31640536]
16. Marcos-Marcos J, Nardini K, Briones-Vozmediano E, Vives-Cases C. Listening to stakeholders in the prevention of gender-based violence among young people in Spain: a qualitative study from the positivMasc project. BMC Womens Health 2023 Jul 26;23(1):393 [FREE Full text] [doi: 10.1186/s12905-023-02545-3] [Medline: 37496067]
17. Casey E, Carlson J, Two Bulls S, Yager A. Gender transformative approaches to engaging men in gender-based violence prevention: a review and conceptual model. Trauma Violence Abuse 2018 Apr;19(2):231-246. [doi: 10.1177/1524838016650191] [Medline: 27197533]
18. Doyle K, Levtov RG, Karamage E, Rakshit D, Kazimbaya S, Sayinzoga F, et al. Long-term impacts of the Bandebereho programme on violence against women and children, maternal health-seeking, and couple relations in Rwanda: a six-year follow-up of a randomised controlled trial. EClinicalMedicine 2023 Sep 26;64:102233 [FREE Full text] [doi: 10.1016/j.eclinm.2023.102233] [Medline: 37781160]
19. van den Berg W, Hendricks L, Hatcher A, Peacock D, Godana P, Dworkin S. 'One man can': shifts in fatherhood beliefs and parenting practices following a gender-transformative programme in Eastern Cape, South Africa. Gend Dev 2013 Feb 01;21(1):111-125 [FREE Full text] [doi: 10.1080/13552074.2013.769775] [Medline: 24000276]
20. Ruane-McAteer E, Gillespie K, Amin A, Aventin Á, Robinson M, Hanratty J, et al. Gender-transformative programming with men and boys to improve sexual and reproductive health and rights: a systematic review of intervention studies. BMJ Glob Health 2020 Oct;5(10):e002997 [FREE Full text] [doi: 10.1136/bmjgh-2020-002997] [Medline: 33051283]
21. WHO guidelines on parenting interventions to prevent maltreatment and enhance parent–child relationships with children aged 0–17 years. World Health Organization. 2023 Feb 10. URL: https://www.who.int/publications/i/item/9789240065505 [accessed 2023-03-09]
22. Backhaus S, Gardner F, Schafer M, Melendez-Torres GJ, Knerr W, Lachman JM. World Health Organization guidelines on parenting interventions to prevent maltreatment and enhance parent–child relationships with children aged 0-17 Years: report of the systematic reviews of evidence. Center for Evidence-Based Intervention, University of Oxford. 2023. URL: https://pure.uva.nl/ws/files/192859093/systematic_reviews-for-the-who-parenting-guideline-jan-27th-2023.pdf [accessed 2024-12-10]
23. Doyle K, Levtov RG, Barker G, Bastian GG, Bingenheimer JB, Kazimbaya S, et al. Gender-transformative Bandebereho couples' intervention to promote male engagement in reproductive and maternal health and violence prevention in Rwanda: findings from a randomized controlled trial. PLoS One 2018 Apr 4;13(4):e0192756 [FREE Full text] [doi: 10.1371/journal.pone.0192756] [Medline: 29617375]

24. Sanders MR, Divan G, Singhal M, Turner KM, Velleman R, Michelson D, et al. Scaling up parenting interventions is critical for attaining the sustainable development goals. Child Psychiatry Hum Dev 2022 Oct;53(5):941-952 [FREE Full text] [doi: 10.1007/s10578-021-01171-0] [Medline: 33948778]
25. Harris M, Andrews K, Gonzalez A, Prime H, Atkinson L. Technology-assisted parenting interventions for families experiencing social disadvantage: a meta-analysis. Prev Sci 2020 Jul;21(5):714-727. [doi: 10.1007/s11121-020-01128-0] [Medline: 32415543]
26. Cluver LD, Meinck F, Steinert JI, Shenderovich Y, Doubt J, Herrero Romero R, et al. Parenting for lifelong health: a pragmatic cluster randomised controlled trial of a non-commercialised parenting programme for adolescents and their families in South Africa. BMJ Glob Health 2018 Jan 31;3(1):e000539 [FREE Full text] [doi: 10.1136/bmjgh-2017-000539] [Medline: 29564157]
27. Ward CL, Wessels IM, Lachman JM, Hutchings J, Cluver LD, Kassanjee R, et al. Parenting for lifelong health for young children: a randomized controlled trial of a parenting program in South Africa to prevent harsh parenting and child conduct problems. J Child Psychol Psychiatry 2020 Apr;61(4):503-512 [FREE Full text] [doi: 10.1111/jcpp.13129] [Medline: 31535371]
28. Schafer M, Lachman JM, Gardner F, Zinser P, Calderon F, Han Q, et al. Integrating intimate partner violence prevention content into a digital parenting chatbot intervention during COVID-19: intervention development and remote data collection. BMC Public Health 2023 Sep 04;23(1):1708 [FREE Full text] [doi: 10.1186/s12889-023-16649-w] [Medline: 37667352]
29. Shai NJ, Sikweyiya Y. Programmes for change: addressing sexual and intimate partner violence in South Africa. S Afr Crime Q 2015 Mar;51:31-41. [doi: 10.4314/sacq.v51i1.4]
30. Sere Y, Roman NV, Ruiter RA. Coping with the experiences of intimate partner violence among South African women: systematic review and meta-synthesis. Front Psychiatry 2021 May 26;12:655130 [FREE Full text] [doi: 10.3389/fpsyt.2021.655130] [Medline: 34122178]
31. Gordon C. Intimate partner violence is everyone's problem, but how should we approach it in a clinical setting? S Afr Med J 2016 Sep 10;106(10):962-965. [doi: 10.7196/samj.2016.v106i10.11408]
32. Violence against women. Organisation for Economic Co-operation and Development. URL: https://www.oecd.org/en/data/indicators/violence-against-women.html [accessed 2024-12-10]
33. Jamaica gender landscape. World Bank Group. URL: https://documents1.worldbank.org/curated/en/099106107012217020/pdf/IDU1101a2e811893114b471a77e16578ca7db7a4.pdf [accessed 2024-12-10]
34. Phaswana ED. Women, gender and race in post-apartheid South Africa. In: Yacob-Haliso O, Falola T, editors. The Palgrave Handbook of African Women's Studies. Cham, Switzerland: Palgrave Macmillan; Feb 09, 2021.
35. Thabane L, Cambon L, Potvin L, Pommier J, Kivits J, Minary L, et al. Population health intervention research: what is the place for pilot studies? Trials 2019 May 30;20(1):309 [FREE Full text] [doi: 10.1186/s13063-019-3422-4] [Medline: 31146768]
36. Kaehler LA, Jacobs M, Jones DJ. Distilling common history and practice elements to inform dissemination: Hanf-model BPT programs as an example. Clin Child Fam Psychol Rev 2016 Sep;19(3):236-258 [FREE Full text] [doi: 10.1007/s10567-016-0210-5] [Medline: 27389606]
37. Ashburn K, Kerner B, Ojamuge D, Lundgren R. Evaluation of the responsible, engaged, and loving (REAL) fathers initiative on physical child punishment and intimate partner violence in northern Uganda. Prev Sci 2017 Oct;18(7):854-864 [FREE Full text] [doi: 10.1007/s11121-016-0713-9] [Medline: 27738782]
38. Kapiga S, Harvey S, Mshana G, Hansen CH, Mtolela GJ, Madaha F, et al. A social empowerment intervention to prevent intimate partner violence against women in a microfinance scheme in Tanzania: findings from the MAISHA cluster randomised controlled trial. Lancet Glob Health 2019 Oct;7(10):e1423-e1434 [FREE Full text] [doi: 10.1016/S2214-109X(19)30316-X] [Medline: 31537372]
39. Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, et al. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. BMJ 2016 Oct 24;355:i5239 [FREE Full text] [doi: 10.1136/bmj.i5239] [Medline: 27777223]
40. Guidelines for producing statistics on violence against women. United Nations. 2014. URL: https://unstats.un.org/unsd/gender/docs/guidelines_statistics_vaw.pdf [accessed 2024-12-10]
41. Gugerty MK, Karlan D. The Goldilocks Challenge: Right-Fit Evidence for the Social Sector. Oxford, UK: Oxford University Press; May 24, 2018.
42. Breitenstein SM, Gross D, Christophersen R. Digital delivery methods of parenting training interventions: a systematic review. Worldviews Evid Based Nurs 2014 Jun;11(3):168-176. [doi: 10.1111/wvn.12040] [Medline: 24842341]
43. Collins DA, Tully LA, Piotrowska PJ, Hawes DJ, Dadds MR. Perspectives on ParentWorks: learnings from the development and national roll-out of a self-directed online parenting intervention. Internet Interv 2019 Mar;15:52-59 [FREE Full text] [doi: 10.1016/j.invent.2018.12.002] [Medline: 30656140]
44. UNICEF's open source approach to innovation. UNICEF Innovation. 2020. URL: https://www.unicef.org./innovation/stories/unicefs-open-source-approach-innovation [accessed 2024-12-08]
45. García-Moreno C, Jansen HA, Watts C, Ellsberg M, Heise L. WHO multi-country study on women's health and domestic violence against women: summary report of initial results on prevalence, health outcomes and women's responses. World

Health Organization. 2005. URL: https://iris.who.int/bitstream/handle/10665/43310/9241593512_eng.pdf?sequence=1 [accessed 2024-12-10]
46. Meinck F, Boyes ME, Cluver L, Ward CL, Schmidt P, DeStone S, et al. Adaptation and psychometric properties of the ISPCAN Child Abuse Screening Tool for use in trials (ICAST-Trial) among South African adolescents and their primary caregivers. Child Abuse Negl 2018 Aug;82:45-58 [FREE Full text] [doi: 10.1016/j.chiabu.2018.05.022] [Medline: 29860107]
47. Ebert C, Steinert JI. Prevalence and risk factors of violence against women and children during COVID-19, Germany. Bull World Health Organ 2021 Jun 01;99(6):429-438 [FREE Full text] [doi: 10.2471/BLT.20.270983] [Medline: 34108753]
48. McNaughton Reyes HL, Maman S, Chen MS, Groves AK, Moodley D. Patterns of intimate partner violence victimization among South African women and their relation to emotional distress during pregnancy and postpartum. J Interpers Violence 2021 Apr 10;36(7-8):NP4230-NP4249 [FREE Full text] [doi: 10.1177/0886260518786738] [Medline: 29991312]
49. Abramsky T, Kapinga I, Mshana G, Lees S, Hansen CH, Hashim R, et al. Couples data from north-western Tanzania: insights from a survey of male partners of women enrolled in the MAISHA cluster randomized trial of an intimate partner violence prevention intervention. PLoS One 2020 Oct 2;15(10):e0240112 [FREE Full text] [doi: 10.1371/journal.pone.0240112] [Medline: 33006997]
50. Lachenbruch PA. McNemar test. In: Wiley StatsRef: Statistics Reference Online. Hoboken, NJ: John Wiley & Sons; 2014.
51. Trajman A, Luiz RR. McNemar chi2 test revisited: comparing sensitivity and specificity of diagnostic examinations. Scand J Clin Lab Invest 2008;68(1):77-80. [doi: 10.1080/00365510701666031] [Medline: 18224558]
52. Ethical and safety recommendations for intervention research on violence against women. World Health Organization. 2016 Nov 29. URL: https://www.who.int/publications/i/item/9789241510189 [accessed 2024-12-10]
53. COVID-19 Risk Communication and Community Engagement (RCCE) guidance note on quantitative data collection. United Nations International Children's Fund. 2020. URL: https://www.corecommitments.unicef.org/kp/rcce-covid19---quantitative-data-collection-guidance-note.pdf [accessed 2024-12-01]
54. Leung L, Miedema S, Warner X, Homan S, Fulu E. Making feminism count: integrating feminist research principles in large-scale quantitative research on violence against women and girls. Gend Dev 2019 Oct 29;27(3):427-447. [doi: 10.1080/13552074.2019.1668142]
55. Jamieson MK, Govaart GH, Pownall M. Reflexivity in quantitative research: a rationale and beginner's guide. Soc Personal Psychol Compass 2023 Feb;17(4):e12735. [doi: 10.1111/spc3.12735]
56. Richardson BP, van der Linde J, Pillay B, Swanepoel DW. Do text messages about health and development in young children affect caregiver behaviour and child outcomes? A systematic review. Health Educ J 2020 Oct 21;80(2):225-237. [doi: 10.1177/0017896920965992]
57. Dadds MR, Sicouri G, Piotrowska PJ, Collins DA, Hawes DJ, Moul C, et al. Keeping parents involved: predicting attrition in a self-directed, online program for childhood conduct problems. J Clin Child Adolesc Psychol 2019 Aug 01;48(6):881-893. [doi: 10.1080/15374416.2018.1485109] [Medline: 30067388]
58. Hall CM, Bierman KL. Technology-assisted interventions for parents of young children: emerging practices, current research, and future directions. Early Child Res Q 2015;33:21-32 [FREE Full text] [doi: 10.1016/j.ecresq.2015.05.003] [Medline: 27773964]
59. Guillot-Wright SP, Lu Y, Torres ED, Le VD, Hall HR, Temple JR. Design and feasibility of a school-based text message campaign to promote healthy relationships. School Ment Health 2018 Feb 9;10(4):428-436. [doi: 10.1007/s12310-018-9255-6]
60. Fletcher R, Kay-Lambkin F, May C, Oldmeadow C, Attia J, Leigh L. Supporting men through their transition to fatherhood with messages delivered to their smartphones: a feasibility study of SMS4dads. BMC Public Health 2017 Dec 13;17(1):953 [FREE Full text] [doi: 10.1186/s12889-017-4978-0] [Medline: 29237439]
61. Lechowicz ME, Jiang Y, Tully LA, Burn MT, Collins DA, Hawes DJ, et al. Enhancing father engagement in parenting programs: translating research into practice recommendations. Aus Psychol 2020 Nov 12;54(2):83-89. [doi: 10.1111/ap.12361]
62. Guthridge M, Kirkman M, Penovic T, Giummarra MJ. Promoting gender equality: a systematic review of interventions. Soc Just Res 2022 Sep 01;35(3):318-343. [doi: 10.1007/s11211-022-00398-z]
63. Yaremych HE, Persky S. Recruiting fathers for parenting research: an evaluation of eight recruitment methods and an exploration of fathers' motivations for participation. Parent Sci Pract 2023 Feb 25;23(1):1-32 [FREE Full text] [doi: 10.1080/15295192.2022.2036940] [Medline: 37346458]
64. Jeong J, Sullivan EF, McCann JK, McCoy DC, Yousafzai AK. Implementation characteristics of father-inclusive interventions in low- and middle-income countries: a systematic review. Ann N Y Acad Sci 2023 Feb;1520(1):34-52 [FREE Full text] [doi: 10.1111/nyas.14941] [Medline: 36482863]
65. 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]
66. Heckman BW, Mathew AR, Carpenter MJ. Treatment burden and treatment fatigue as barriers to health. Curr Opin Psychol 2015 Oct 01;5:31-36 [FREE Full text] [doi: 10.1016/j.copsyc.2015.03.004] [Medline: 26086031]
67. Hanno EC. Nudging early educators' knowledge, beliefs, and practices: an embedded randomized controlled trial of text message supports. Early Educ Dev 2022 Mar 20;34(2):530-550. [doi: 10.1080/10409289.2022.2053476]

68. Cooper H, Nadzri FZ, Vyas S, Juhari R, Ismail N, Arshat Z, et al. A hybrid digital parenting program delivered within the Malaysian preschool system: protocol for a feasibility study of a small-scale factorial cluster randomized trial. JMIR Res Protoc 2024 Apr 26;13:e55491 [FREE Full text] [doi: 10.2196/55491] [Medline: 38669679]
69. UNICEF Innocenti - Global Office of Research and Foresight, Prevention Collaborative, Equimundo. Parenting programmes to reduce violence against children and women. Why it is important. Brief 1.: UNICEF Innocenti; 2023. URL: https://www.unicef.org/innocenti/media/3456/file/UNICEF-Innocenti-Parenting-Programmes-Reduce-Violence-Against-Children-Women-Brief-1-2024-EN.pdf
70. De Filippo A, Bellatin P, Tietz N, Grant E, Whitefield A, Nkopane P, et al. Effects of digital chatbot on gender attitudes and exposure to intimate partner violence among young women in South Africa. PLOS Digit Health 2023 Oct 16;2(10):e0000358 [FREE Full text] [doi: 10.1371/journal.pdig.0000358] [Medline: 37844088]
71. Kågesten A, Gibbs S, Blum RW, Moreau C, Chandra-Mouli V, Herbert A, et al. Understanding factors that shape gender attitudes in early adolescence globally: a mixed-methods systematic review. PLoS One 2016 Jun 24;11(6):e0157805 [FREE Full text] [doi: 10.1371/journal.pone.0157805] [Medline: 27341206]
72. Falb KL, Asghar K, Blackwell A, Baseme S, Nyanguba M, Roth D, et al. Improving family functioning and reducing violence in the home in North Kivu, Democratic Republic of Congo: a pilot cluster-randomised controlled trial of Safe at Home. BMJ Open 2023 Mar 06;13(3):e065759 [FREE Full text] [doi: 10.1136/bmjopen-2022-065759] [Medline: 36878658]
73. Halim N, Steven Mzilangwe E, Reich N, Badi L, Simmons E, Servidone M, et al. Together to end violence against women in Tanzania: results of a pilot cluster randomized controlled trial to evaluate preliminary effectiveness of interpersonal and community level interventions to address intimate partner violence. Glob Public Health 2019 Dec 13;14(12):1653-1668 [FREE Full text] [doi: 10.1080/17441692.2019.1609062] [Medline: 31084264]
74. Gender action plan, 2022-2025: a vision for lasting, transformative change. United Nations Children's Fund. URL: https://www.unicef.org/gender-equality/gender-action-plan-2022-2025
75. Collins LM. Optimization of Behavioral, Biobehavioral, and Biomedical Interventions: The Multiphase Optimization Strategy (MOST). Cham, Switzerland: Springer; 2018.
76. Starmann E, Collumbien M, Kyegombe N, Devries K, Michau L, Musuya T, et al. Exploring couples' processes of change in the context of SASA!, a violence against women and HIV prevention intervention in Uganda. Prev Sci 2017 Feb 29;18(2):233-244 [FREE Full text] [doi: 10.1007/s11121-016-0716-6] [Medline: 27682273]
77. van Aar J, Leijten P, Orobio de Castro B, Overbeek G. Sustained, fade-out or sleeper effects? A systematic review and meta-analysis of parenting interventions for disruptive child behavior. Clin Psychol Rev 2017 Feb;51:153-163. [doi: 10.1016/j.cpr.2016.11.006] [Medline: 27930935]
78. Broadband statistics. Organisation for Economic Co-operation and Development. URL: https://www.oecd.org/en/topics/sub-issues/broadband-statistics.html [accessed 2023-10-31]
79. Awah I, Green O, Baerecke L, Janowski R, Klapwijk J, Chetty AN, et al. 'It provides practical tips, practical solutions!': acceptability, usability, and satisfaction of a digital parenting intervention across African countries. Psychol Health Med 2022 Aug 18;27(sup1):107-123. [doi: 10.1080/13548506.2022.2113106] [Medline: 35980251]
80. Janowski R, Green O, Shenderovich Y, Stern D, Clements L, Wamoyi J, et al. Optimising engagement in a digital parenting intervention to prevent violence against adolescents in Tanzania: protocol for a cluster randomised factorial trial. BMC Public Health 2023 Jun 23;23(1):1224 [FREE Full text] [doi: 10.1186/s12889-023-15989-x] [Medline: 37353844]
81. Cullen C. Method matters: the underreporting of intimate partner violence. World Bank Econ Rev 2023 Feb;37(1):49-73. [doi: 10.1093/wber/lhac022]
82. Visschers J, Jaspaert E, Vervaeke G. Social desirability in intimate partner violence and relationship satisfaction reports: an exploratory analysis. J Interpers Violence 2017 May 09;32(9):1401-1420. [doi: 10.1177/0886260515588922] [Medline: 26058978]
83. McGirr S, Torres J, Heany J, Brandon H, Tarry C, Robinson C. Lessons learned on recruiting and retaining young fathers in a parenting and repeat pregnancy prevention program. Matern Child Health J 2020 Sep 20;24(Suppl 2):183-190 [FREE Full text] [doi: 10.1007/s10995-020-02956-w] [Medline: 32564249]
84. Bott S, Guedes A, Subrahmanian R, Chowdhury S, Rumble L, Daban RA. Gender dimensions of violence against children and adolescents. United Nations Children's Fund (UNICEF). 2020 Apr. URL: https://www.unicef.org/media/93986/file/Child-Protection-Gender-Dimensions-of-VACAG-2021.pdf [accessed 2024-12-10]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
IPV: intimate partner violence
ODK: Open Data Kit
RCT: randomized controlled trial
UNICEF: United Nations Children’s Fund
VAC: violence against children
VAW: violence against women

Edited by A Mavragani; submitted 20.03.24; peer-reviewed by M Wang, S Willan, RG Levtov; comments to author 25.06.24; revised version received 13.10.24; accepted 30.10.24; published 03.01.25.

Please cite as:
Schafer M, Lachman J, Zinser P, Calderón Alfaro FA, Han Q, Facciola C, Clements L, Gardner F, Haupt Ronnie G, Sheil R
A Digital Parenting Intervention With Intimate Partner Violence Prevention Content: Quantitative Pre-Post Pilot Study
JMIR Form Res 2025;9:e58611
URL: https://formative.jmir.org/2025/1/e58611
doi:10.2196/58611
PMID:

Short Paper

Validation of Sleep Measurements of an Actigraphy Watch: Instrument Validation Study

Mari Waki[1*]; Ryohei Nakada[1*]; Kayo Waki[1]; Yuki Ban[1]; Ryo Suzuki[2]; Toshimasa Yamauchi[3]; Masaomi Nangaku[4]; Kazuhiko Ohe[1]

[1]Department of Biomedical Informatics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

[2]Department of Diabetes, Metabolism and Endocrinology, Tokyo Medical University, Tokyo, Japan

[3]Department of Diabetes and Metabolic Diseases, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

[4]Division of Nephrology and Endocrinology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

[*]these authors contributed equally

Corresponding Author:
Kayo Waki
Department of Biomedical Informatics
Graduate School of Medicine
The University of Tokyo
7-chōme-3-1 Hongō
Bunkyo City
Tokyo, 113-8654
Japan
Phone: 81 03 5800 9129
Fax: 81 03 5800 9173
Email: kwaki-tky@m.u-tokyo.ac.jp

Abstract

Background: The iAide2 (Tokai) physical activity monitoring system includes diverse measurements and wireless features useful to researchers. The iAide2's sleep measurement capabilities have not been compared to validated sleep measurement standards in any published work.

Objective: We aimed to assess the iAide2's sleep duration and total sleep time (TST) measurement performance and perform calibration if needed.

Methods: We performed free-living sleep monitoring in 6 convenience-sampled participants without known sleep disorders recruited from within the Waki DTx Laboratory at the Graduate School of Medicine, University of Tokyo. To assess free-living sleep, we validated the iAide2 against a second actigraph that was previously validated against polysomnography, the MotionWatch 8 (MW8; CamNtech Ltd). The participants wore both devices on the nondominant arm, with the MW8 closest to the hand, all day except when bathing. The MW8 and iAide2 assessments both used the MW8 EVENT-marker button to record bedtime and risetime. For the MW8, MotionWare Software (version 1.4.20; CamNtech Ltd) provided TST, and we calculated sleep duration from the sleep onset and sleep offset provided by the software. We used a similar process with the iAide2, using iAide2 software (version 7.0). We analyzed 64 nights and evaluated the agreement between the iAide2 and the MW8 for sleep duration and TST based on intraclass correlation coefficients (ICCs).

Results: The absolute ICCs (2-way mixed effects, absolute agreement, single measurement) for sleep duration (0.69, 95% CI –0.07 to 0.91) and TST (0.56, 95% CI –0.07 to 0.82) were moderate. The consistency ICC (2-way mixed effects, consistency, single measurement) was excellent for sleep duration (0.91, 95% CI 0.86-0.95) and moderate for TST (0.78, 95% CI 0.67-0.86). We determined a simple calibration approach. After calibration, the ICCs improved to 0.96 (95% CI 0.94-0.98) for sleep duration and 0.82 (95% CI 0.71-0.88) for TST. The results were not sensitive to the specific participants included, with an ICC range of 0.96-0.97 for sleep duration and 0.79-0.87 for TST when applying our calibration equation to data removing one participant at a time and 0.96-0.97 for sleep duration and 0.79-0.86 for TST when recalibrating while removing one participant at a time.

Conclusions: The measurement errors of the uncalibrated iAide2 for both sleep duration and TST seem too large for them to be useful as absolute measurements, though they could be useful as relative measurements. The measurement errors after calibration

are low, and the calibration approach is general and robust, validating the use of iAide2's sleep measurement functions alongside its other features in physical activity research.

(JMIR Form Res 2025;9:e63529) doi:10.2196/63529

KEYWORDS

actigraphy; sleep; Motion Watch 8; iAide2; total sleep time

Introduction

Background

The iAide2 (Tokai) is a physical activity monitoring system that measures intensity levels, step counts, temperature, and pulse rate [1,2] to assess activity and sleep, with minute-by-minute assessment of wake/sleep [3,4]. It is based on TDK's Silmee W20/W22 actigraphs [4-6] used in previous research [5,7-9]. The iAide2 supports long trials (via rechargeability and off-watch data storage) and timely behavior feedback (via a wireless EVENT button and near–real-time access to data). Its sleep measurements are not well validated; Kimura et al [7] compared its predecessor, the W20, to video-monitoring sleep assessment (which is itself unvalidated), and reported 5 sleep duration values, with detailed methodology not given [7].

Actigraphs are commonly validated against polysomnography (PSG) or another actigraph. PSG, though imperfect, is the gold standard of sleep measurement [10]. PSG's complexity requires sleep in a laboratory [11], which is costly and unrepresentative of free-living sleep. Watch-based actigraphy can efficiently collect multiple nights' data in a natural environment [12,13].

Objective

We assessed iAide2-measured sleep duration and total sleep time (TST) in participants without sleep disorders and added calibration when needed.

Methods

Approach

To assess free-living sleep, we validated iAide2 against a second actigraph, the MotionWatch 8 (MW8; CamNtech Ltd) [14], using methods similar to other actigraph-to-actigraph comparisons [11,15-17]. The MW8 is validated against PSG [18-20], with high correlations (r=0.920) and no significant bias (7.0 minutes) for TST [19]. Researchers have used the MW8 to assess sleep in people without sleep disorders [21-23]. We used validated MW8 settings (mode 1, 30-second epochs, and a threshold of 20 [24,25]).

Data Collection

We collected data from a convenience sample of members of the Waki DTx Laboratory at the Graduate School of Medicine, University of Tokyo. The laboratory is investigating lifestyle interventions to improve sleep and exercise and identified the iAide2 as a candidate measurement device. This validation study assessed the accuracy of iAide2 sleep monitoring in 6 individuals without known sleep disorders. They wore both devices on the nondominant arm, with the MW8 closest to the hand, all day except when bathing.

The participants used the MW8 EVENT-marker button to record bedtime (when they began trying to sleep) and risetime (when they stopped trying to sleep). The duration between these times is time in bed (TIB). Our MW8 and iAide2 assessments both used these nonactigraphy measurements.

MotionWare (version 1.4.20; CamNtech Ltd) provided sleep onset ("Fell Asleep" in MotionWare, the earliest time after bedtime when the participant was asleep), sleep offset ("Woke Up," the latest time prior to risetime when the participant transitioned from asleep to awake), and TST (the sum of all minutes of sleep between sleep onset and offset) [25]. From these measurements, we derived sleep onset latency (the duration between bedtime and sleep onset), sleep offset latency (the duration between sleep offset and risetime), and sleep duration (the duration between sleep onset and sleep offset).

We used a similar process with the iAide2, using iAide2 software (version 7.0). We defined sleep onset as the time of the first sleep after bedtime. Similarly, we defined sleep offset as 1 minute after the time of final sleep prior to risetime. We defined TST as the sum of minutes asleep within the sleep duration period.

Statistical Analysis

We compared measurements using intraclass correlation coefficients (ICCs; 2-way mixed effects, absolute agreement, single measurement, unless otherwise specified) [26,27]. We used the *irr* package 0.84.1 in R for point estimation of ICC and calculating 95% CIs, and R (version 4.2.1) for calibration.

Ethical Considerations

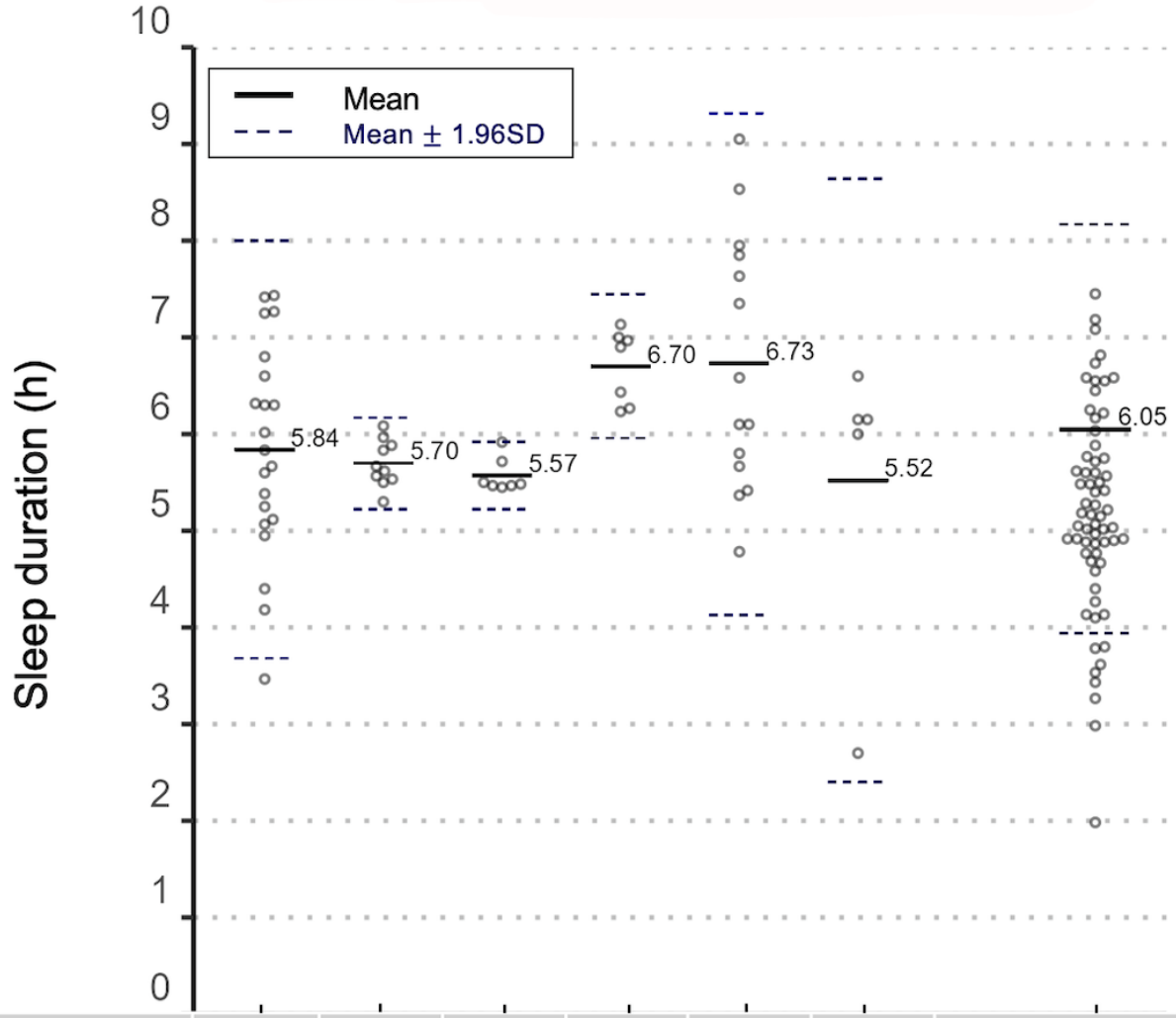
The University of Tokyo School of Medicine institutional review board (2024137NI) approved this study, which abided by the Declaration of Helsinki. Participants provided verbal consent for data collection upon recruitment. All participants provided written consent for the use of their data. Participants' data were anonymized. No compensation was provided.

Results

Participants

We recruited 6 participants (aged 24-63 years) and recorded data for 64 nights (Figure 1). The number of nights per participant varied from 5 to 21. The average nightly sleep duration via MW8 per participant varied from 5.52 to 6.73 hours, and the participants differed in the regularity of their sleep durations.

Figure 1. Participants' information and sleep duration measured using the MotionWatch 8 actigraph (MW8; 6 participants, 64 nights).



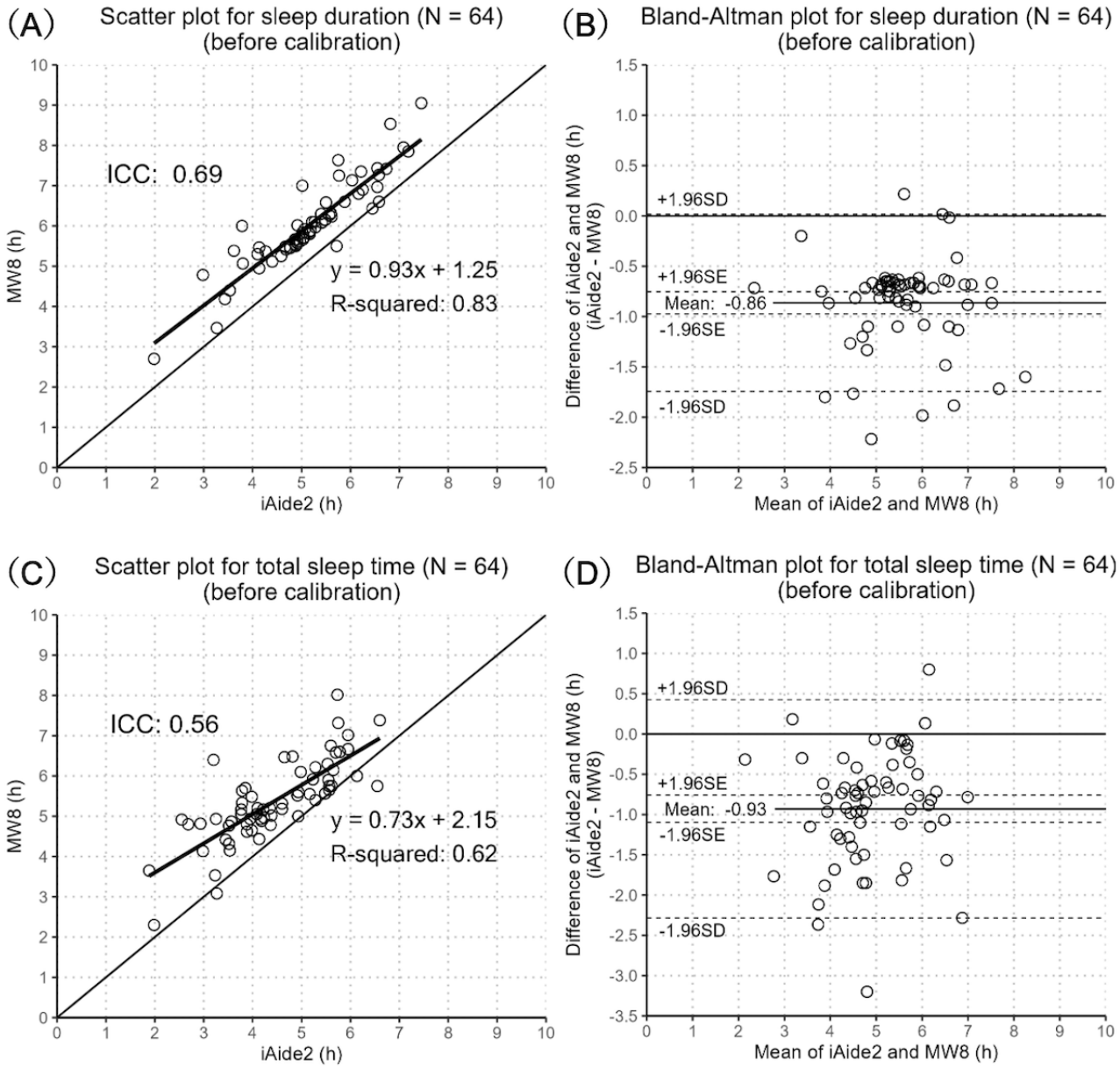
ID	4	9	10	7	11	12	Total
Nights, n	21	10	7	7	14	5	64
Age (yrs)	24	50	54	63	55	35	
Sex	M	F	F	M	M	F	
Sleep duration (h), mean (95%CI)	5.84 [5.30-6.34]	5.70 [5.52-5.87]	5.57 [5.41-5.74]	6.70 [6.35-7.06]	6.73 [5.96-7.49]	5.52 [3.54-7.50]	6.05 [5.78-6.32]

Measurements

The iAide2 underestimated duration relative to the MW8 and showed moderate ICC of 0.69 (95% CI –0.07 to 0.91; Figure 2A) [26]. Relative performance was excellent (ICC=0.91, 95% CI 0.86-0.95; 2-way mixed effects, consistency, single measurement) [26]. The mean, maximum, and minimum differences between iAide2 and MW8 were –0.86, 0.22, and –2.22 hours, respectively (Figure 2B).

Similarly, the iAide2 underestimated TST and showed moderate ICC of 0.56 (95% CI –0.07 to 0.82; Figure 2C) and a good consistency ICC (2-way mixed effects, consistency, single measurement) of 0.78 (95% CI 0.67-0.86). The mean, maximum, and minimum differences between iAide2 and MW8 were –0.93, 0.80, and –3.20 hours, respectively (Figure 2D).

Figure 2. Scatter plot and Bland-Altman plot for measured sleep duration and total sleep time using iAide2 and the MotionWatch 8 actigraph (MW8; 64 nights).



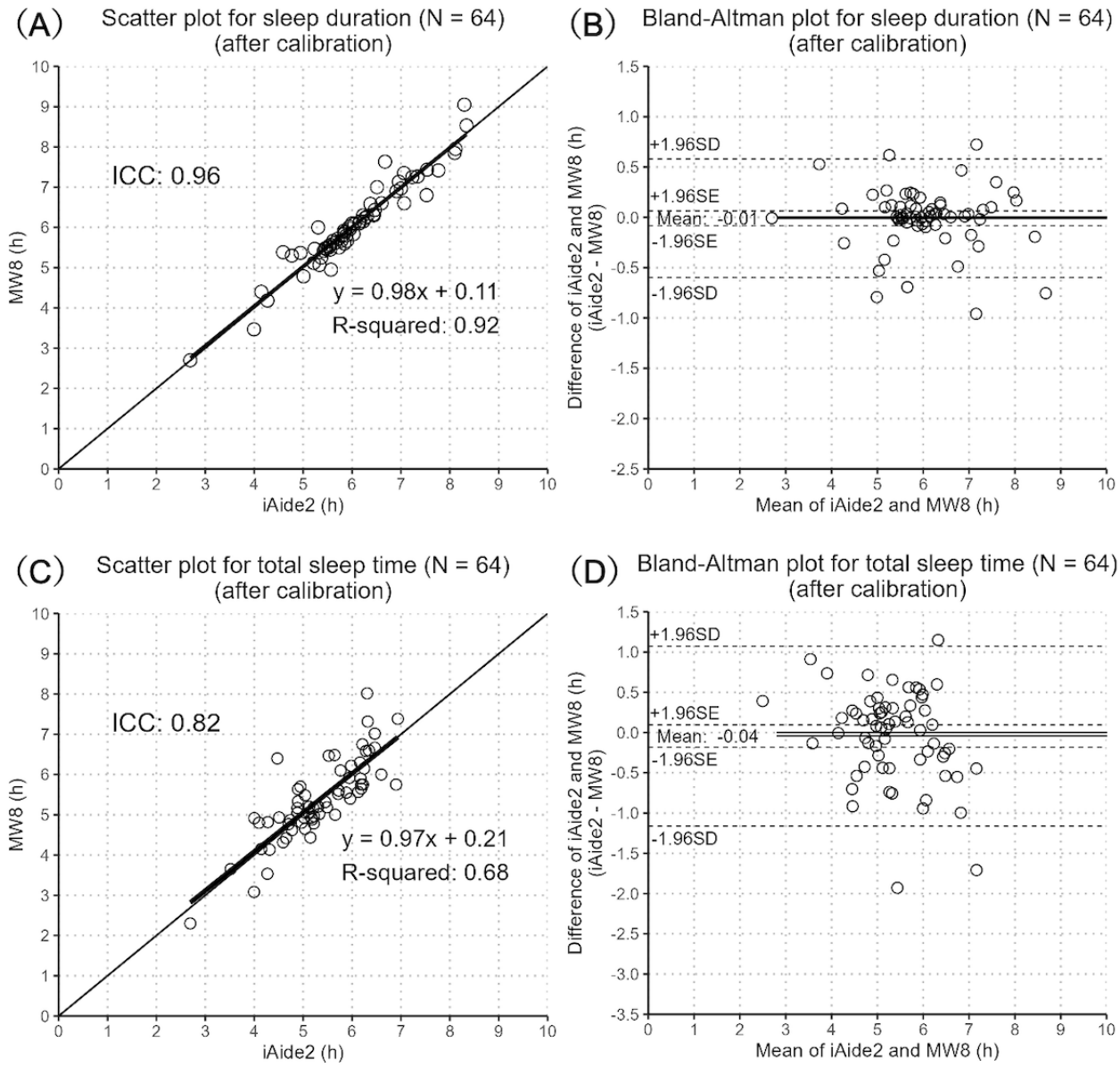
Calibration

Our iAide2 data showed very small sleep offset latencies (mean 0.03 hours) but significant overestimates of sleep onset latency, so we corrected the sleep duration by the difference between the iAide2-measured sleep onset latency and an iAide2 sleep onset latency calibrated to the MW8 measurements via ordinary least squares (OLS). We added a small (<5 minute) bias correction, selected to optimize ICC, to account for small offset latency biases. To prevent physically impossible results, we limited the calibrated duration to be no longer than the TIB. The resulting calibration equation is D' = min {0.59S + D + 0.26, TIB} (hours), where D is the measured duration, S is the measured sleep onset latency, and D' is the calibrated duration.

For TST, we used OLS to calibrate the measured TST value, limiting the result to be not greater than the calibrated sleep duration. The resulting equation is T' = min {0.73T + 2.15, D'} (hours), where T is the measured TST and T' is the calibrated TST.

Calibration greatly improved accuracy. Sleep duration had excellent ICC (0.96, 95% CI 0.94-0.98; Figure 3A). The mean difference almost disappeared (–0.01 hours) and the difference between the maximum (0.72 hours) and minimum (–0.96 hours) values decreased (Figure 3B). TST similarly had greatly improved accuracy, with good ICC (0.82, 95% CI 0.71-0.88; Figure 3C). The mean difference became almost zero (–0.04 hours) and the difference between the maximum (1.15 hours) and minimum (–1.93 hours) values decreased (Figure 3D).

Figure 3. Scatter plot and Bland-Altman plot for calibrated sleep duration and total sleep time using iAide2 and the MotionWatch 8 actigraph (MW8; 64 nights).



Because our sample size is small with a varying number of nights per participant, we conducted a sensitivity analysis to determine how the ICC agreement rate changes based on the included participants. First, to show that our ICCs are not sensitive to the exact individuals, we performed trials excluding the data of one participant at a time, recalibrating for each trial. The calibrated ICC range was 0.96-0.97 for sleep duration, effectively unchanged from 0.96 in the baseline analysis, and 0.79-0.86 for TST, similar to 0.82 in the baseline analysis, demonstrating that our results are not highly influenced by the variability in number of nights per participant. To demonstrate the generality of our calibration equations, we applied the calibration derived from all participants' data to trials excluding one participant at a time. The ICC range was 0.96-0.97 for sleep duration and 0.79-0.87 for TST, again similar to baseline analysis values, showing that our calibration has high generality and is not highly dependent on any one participant's data.

Discussion

Principal Findings

The absolute accuracies of uncalibrated iAide2 measurements for sleep duration (ICC=0.69) and TST (ICC=0.56) are relatively poor. Uncalibrated data could be useful in applications needing relative measurements, with good values for the consistency version of the ICC (2-way mixed effects, consistency, single measurement) of 0.91 for sleep duration and 0.78 for TST.

The calibrated measurements have excellent absolute accuracy. The ICCs for sleep duration (0.96) and TST (0.82) are comparable to or better than those reported previously (0.67 and 0.68 for sleep duration and 0.77 and 0.84 for TST [28,29]). Sensitivity analysis revealed essentially unchanged performance when applying this calibration to data while removing one participant at a time and when recalibrating while removing one participant at a time. The calibration methods are simple

and have good statistical bases. The calibrated iAide2 measurements of sleep duration and TST seem usable in a wide range of applications.

Limitations

The MW8, our validation standard, has its own errors, with a reported underestimation of sleep onset latency (mean difference 8.9 minutes) and overestimation of TST (mean difference 47.1 minutes) relative to PSG [30]. Validation against other standards, including PSG, is warranted.

We used more nights (N=64) than some PSG-based validation studies (N=20, 37, and 38) but fewer than actigraph-based research (N=140) [17,18,31,32]. Our convenience sample was small (N=6); however, it included a mix of ages, sex, and sleeping regularity. Our results may not fully apply to other populations; and larger studies are warranted.

Conclusions

The iAide2 has interesting measurement capabilities with convenient features, and validating its sleep measurement accuracy adds to its usefulness. The absolute accuracies are low, but good relative accuracy may be suitable for some applications. Its postcalibration accuracies are excellent and well validated against the MW8.

Acknowledgments

We thank Daniel Lane for manuscript editing and scientific discussions.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request for noncommercial purposes.

Authors' Contributions

KW was responsible for the overall design and evaluation of this study. The model was designed and implemented by YB. Data were collected and organized by YB, RN, and MW. Experiments were conducted by YB, RN, and MW. KW and RS supervised the work. The first draft of the manuscript was written by MW. All authors read and approved the final manuscript.

Conflicts of Interest

None.

References

1. iAide2 (iAide2-W). Pharmaceuticals and Medical Devices Agency. 2021. URL: https://www.info.pmda.go.jp/downfiles/md/PDF/340205/340205_301AGBZX00020000_A_01_02.pdf [accessed 2024-02-26]
2. iAide. Tokai Corporation. URL: https://www.iaide-system.jp/iAide/Pages/Login.aspx [accessed 2024-04-10]
3. Watch-type activity meters that can be used in various businesses Silmee W22. TDK Product Center. URL: https://product.tdk.com/ja/products/biosensor/biosensor/silmee_w22/index.html [accessed 2024-02-26]
4. Miyamoto K, Hashimoto K, Kasaoka M, Kakumu M. Wearable sensors corresponding to various applications in medical/healthcare field. 2020 Presented at: 27th International Workshop on Active-Matrix Flatpanel Displays and Devices (AM-FPD); September 1-4, 2020; Kyoto, Japan. [doi: 10.23919/am-fpd49417.2020.9224474]
5. Detection of atrial fibrillation using a deep learning with wearable pulse wave sensor. 2019 Presented at: APHRS (ASIA Pacific Heart Rhythm Society scientific session); Oct 24-27, 2019; Bangkok, Thailand. [doi: 10.1002/joa3.12267]
6. Biometric Sensors: Business collaboration to commercialize 24-hour remote monitoring system for medical and nursing care using wristband-type wearable devices. TDK Corporation. 2017. URL: https://www.tdk.com/ja/news_center/press/20171002_01.html [accessed 2024-02-26]
7. Kimura N, Aso Y, Yabuuchi K, Ishibashi M, Hori D, Sasaki Y, et al. Modifiable lifestyle factors and cognitive function in older people: a cross-sectional observational study. Front Neurol 2019 Apr 24;10:401 [FREE Full text] [doi: 10.3389/fneur.2019.00401] [Medline: 31068892]
8. Tazawa Y, Liang K, Yoshimura M, Kitazawa M, Kaise Y, Takamiya A, et al. Evaluating depression with multimodal wristband-type wearable device: screening and assessing patient severity utilizing machine-learning. Heliyon 2020 Feb;6(2):e03274 [FREE Full text] [doi: 10.1016/j.heliyon.2020.e03274] [Medline: 32055728]
9. Kadoya Y, Khan MSR, Watanapongvanich S, Binnagan P. Emotional status and productivity: evidence from the Special Economic Zone in Laos. Sustainability 2020 Feb 19;12(4):1544. [doi: 10.3390/su12041544]
10. Tryon WW. Nocturnal activity and sleep assessment. Clin Psychol Rev 1996 Jan;16(3):197-213. [doi: 10.1016/0272-7358(95)00059-3]
11. Asgari Mehrabadi M, Azimi I, Sarhaddi F, Axelin A, Niela-Vilén H, Myllyntausta S, et al. Sleep tracking of a commercially available smart ring and smartwatch against medical-grade actigraphy in everyday settings: instrument validation study. JMIR Mhealth Uhealth 2020 Nov 02;8(10):e20465. [doi: 10.2196/20465] [Medline: 33038869]

12. Fekedulegn D, Andrew ME, Shi M, Violanti JM, Knox S, Innes KE. Actigraphy-based assessment of sleep parameters. Ann Work Expo Health 2020 Apr 30;64(4):350-367 [FREE Full text] [doi: 10.1093/annweh/wxaa007] [Medline: 32053169]
13. Ancoli-Israel S, Cole R, Alessi C, Chambers M, Moorcroft W, Pollak CP. The role of actigraphy in the study of sleep and circadian rhythms. Sleep 2003 May 01;26(3):342-392. [doi: 10.1093/sleep/26.3.342] [Medline: 12749557]
14. MotionWatch 8. CamNtech Ltd. URL: https://www.camntech.com/motionwatch-8/ [accessed 2024-02-26]
15. Morgan PL, Hampton S, Karatziotou A, Zaslona J, Middleton B. Validation of two new activity monitors: motionwatch 8 and pro-diary motion. 2012 Presented at: 21st Congress of the European-Sleep-Research-Society; September 4-8, 2012; Paris, France.
16. Topalidis P, Florea C, Eigl E, Kurapov A, Leon CAB, Schabus M. Evaluation of a low-cost commercial actigraph and its potential use in detecting cultural variations in physical activity and sleep. Sensors (Basel) 2021 May 29;21(11) [FREE Full text] [doi: 10.3390/s21113774] [Medline: 34072347]
17. Shah VV, Brumbach BH, Pearson S, Vasilyev P, King E, Carlson-Kuhta P, et al. Opal actigraphy (activity and sleep) measures compared to ActiGraph: a validation study. Sensors (Basel) 2023 Feb 18;23(4):2296 [FREE Full text] [doi: 10.3390/s23042296] [Medline: 36850896]
18. Elbaz M, Yauy K, Metlaine A, Martoni M, Leger D. Validation of a new actigraph motion watch versus polysomnography on 70 healthy and suspected sleep-disordered subjects. CamNtech Ltd. URL: https://www.camntech.com/Engineering/MotionWare/MotionWatchPSG.pdf [accessed 2023-10-19]
19. Dorffner G, Kemethofer M, Saletu-Zyhlarz G, Simeoni M, Rainer L, Parapatics S, et al. Measuring sleep: a comparison of polysomnography (PSG), wrist actigraphy and a reduced PSG solution. International Society for CNS Clinical Trials and Methodology. URL: https://isctm.org/public_access/Autumn2018/Abstracts/Dorffner-abstract.pdf [accessed 2023-10-17]
20. Yavuz-Kodat E, Reynaud E, Geoffray M, Limousin N, Franco P, Bourgin P, et al. Validity of actigraphy compared to polysomnography for sleep assessment in children with autism spectrum disorder. Front Psychiatry 2019 Aug 2;10:551 [FREE Full text] [doi: 10.3389/fpsyt.2019.00551] [Medline: 31428003]
21. Deantoni M, Reyt M, Baillet M, Dourte M, De Haan S, Lesoinne A, et al. Napping and circadian sleep-wake regulation during healthy aging. Sleep 2024 May 10;47(5):2023 [FREE Full text] [doi: 10.1093/sleep/zsad287] [Medline: 37943833]
22. Illingworth G, Sharman R, Harvey C, Foster RG, Espie CA. The Teensleep study: the effectiveness of a school-based sleep education programme at improving early adolescent sleep. Sleep Med X 2020 Dec;2:100011 [FREE Full text] [doi: 10.1016/j.sleepx.2019.100011] [Medline: 33458648]
23. Nédélec M, Chauvineau M, Guilhem G. On the road to Camarón: the sleep of an ultra-endurance athlete cycling 10,000 km in 24 days. Int J Environ Res Public Health 2022 Apr 09;19(8):4543 [FREE Full text] [doi: 10.3390/ijerph19084543] [Medline: 35457410]
24. The MotionWatch User Guide. CamNtech. 2024. URL: https://www.camntech.com/Products/MotionWatch/The%20MotionWatch%20User%20Guide.pdf [accessed 2024-02-26]
25. MotionWare Software User Guide. CamNtech. 2024. URL: https://www.camntech.com/Products/MotionWatch/The%20MotionWare%20Software%20User%20Guide.pdf [accessed 2024-02-26]
26. 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. [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]
27. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Methods 1996 Mar;1(1):30-46. [doi: 10.1037/1082-989X.1.1.30]
28. Plekhanova T, Rowlands AV, Davies MJ, Hall AP, Yates T, Edwardson CL. Validation of an automated sleep detection algorithm using data from multiple accelerometer brands. J Sleep Res 2023 Jun 31;32(3):e13760. [doi: 10.1111/jsr.13760] [Medline: 36317222]
29. Kainec KA, Caccavaro J, Barnes M, Hoff C, Berlin A, Spencer RMC. Evaluating accuracy in five commercial sleep-tracking devices compared to research-grade actigraphy and polysomnography. Sensors (Basel) 2024 Jan 19;24(2) [FREE Full text] [doi: 10.3390/s24020635] [Medline: 38276327]
30. Ameen MS, Cheung LM, Hauser T, Hahn MA, Schabus M. About the accuracy and problems of consumer devices in the assessment of sleep. Sensors (Basel) 2019 Sep 25;19(19) [FREE Full text] [doi: 10.3390/s19194160] [Medline: 31557952]
31. Kanady JC, Ruoff L, Straus LD, Varbel J, Metzler T, Richards A, et al. Validation of sleep measurement in a multisensor consumer grade wearable device in healthy young adults. J Clin Sleep Med 2020 Jun 15;16(6):917-924 [FREE Full text] [doi: 10.5664/jcsm.8362] [Medline: 32048595]
32. Kosmadopoulos A, Sargent C, Darwent D, Zhou X, Roach GD. Alternatives to polysomnography (PSG): a validation of wrist actigraphy and a partial-PSG system. Behav Res Methods 2014 Dec;46(4):1032-1041. [doi: 10.3758/s13428-013-0438-7] [Medline: 24442593]

Abbreviations

ICC: intraclass correlation coefficient
MW8: MotionWatch 8 actigraph
OLS: ordinary least squares

PSG: polysomnography
TIB: time in bed
TST: total sleep time

Edited by A Mavragani; submitted 23.06.24; peer-reviewed by Y Minami, JE Lee; comments to author 25.10.24; revised version received 11.11.24; accepted 24.11.24; published 06.01.25.

Please cite as:
Waki M, Nakada R, Waki K, Ban Y, Suzuki R, Yamauchi T, Nangaku M, Ohe K
Validation of Sleep Measurements of an Actigraphy Watch: Instrument Validation Study
JMIR Form Res 2025;9:e63529
URL: https://formative.jmir.org/2025/1/e63529
doi:10.2196/63529
PMID:

Original Paper

Differences in eHealth Access, Use, and Perceived Benefit Between Different Socioeconomic Groups in the Dutch Context: Secondary Cross-Sectional Study

Lucille Standaar[1,2,3], BSc, MSc; Lilian van Tuyl[2], MBA, PhD; Anita Suijkerbuijk[1], PhD; Anne Brabers[2], PhD; Roland Friele[2,3], Prof Dr

[1]Department of Population Health and Health Services Research, Centre for Public Health, Healthcare and Society, National Institute for Public Health and the Environment, Bilthoven, Netherlands

[2]Department Organisation and Quality of Care, Netherlands Institute for Health Services Research, Utrecht, Netherlands

[3]Tranzo Scientific Center for Care and Wellbeing, Tilburg University, Tilburg, Netherlands

Corresponding Author:
Lucille Standaar, BSc, MSc
Department of Population Health and Health Services Research
Centre for Public Health, Healthcare and Society
National Institute for Public Health and the Environment
PO Box 1
Bilthoven, 3723 BA
Netherlands
Phone: 31 302743243
Fax: 31 302744407
Email: lucille.standaar@rivm.nl

Abstract

Background: There is a growing concern that digital health care may exacerbate existing health disparities. Digital health care or eHealth encompasses the digital apps that are used in health care. Differences in access, use, and perceived benefits of digital technology among socioeconomic groups are commonly referred to as the digital divide. Current research shows that people in lower socioeconomic positions (SEPs) use eHealth less frequently.

Objective: This study aims to (1) investigate the association between SEP and eHealth access to, use of, and perceived benefit within the adult Dutch population and (2) evaluate disparities in eHealth access, use, and perceived benefit through three socioeconomic variables—education, standardized income, and the socioeconomic status of the neighborhood.

Methods: A secondary analysis was conducted on data from the Nivel Dutch Health Care Consumer Panel (response rate 57%, 849/1500, to assess access to, use of, and perceived benefits from eHealth. These data were collected to monitor eHealth developments in the Netherlands. eHealth was examined through two concepts: (1) eHealth in general and (2) websites, apps, and wearables. Results were stratified into 9 SEP populations based on 3 indicators—education, standardized income, and socioeconomic status level of the neighborhood. Logistic regression analyses were performed to evaluate whether the outcomes varied significantly across different SEP groups. Age was included as a covariate to control for confounding.

Results: This study confirms the association between eHealth and SEP and shows that low SEP respondents have less access (odds ratio [OR] 5.72, 95% CI 3.06-10.72) and use (OR 4.96, 95% CI 2.66-9.24) of eHealth compared to medium or high SEP respondents. Differences were most profound when stratifying for levels of education.

Conclusions: The access to and use of eHealth has a socioeconomic gradient and emphasizes that SEP indicators cannot be used interchangeably to assess eHealth access and use. The results underline the importance of activities and policies aimed at improving eHealth accessibility and usage among low SEP groups to mitigate disparities in health between different socioeconomic groups.

(JMIR Form Res 2025;9:e49585) doi:10.2196/49585

KEYWORDS

eHealth; digital divide; socioeconomic factors; education; income; neighborhood; health disparities; cross-sectional studies; digital health care; health equity; Netherlands

Introduction

Digital health care is expected to provide benefits for health care systems, providers, and patients and is considered a solution to address workforce shortages and rising health care costs [1-4]. Moreover, digital health care is anticipated to enhance the quality of care, stimulate patient self-management, and improve health accessibility and equity [1-4]. Digital health care or eHealth is defined as the digital apps that are used in health care. Health care policies focus on increased use of and dependency on the eHealth apps [4-6]. Concerns have been raised that digital health care may not be equally accessible for all [7-13]. Populations with lower socioeconomic positions (SEPs) are more likely to encounter financial, skill, or cognitive barriers to accessing and using eHealth, such as limited access to devices, limited digital health skills, or limited ability to take the initiative in using eHealth [7,11-16]. Additionally, these populations experience challenges in comprehending and implementing health information and healthy behaviors in daily life [17-21].

The demand for health care services is often higher in low-SEP populations, as people with a low SEP endure more often from chronic illnesses [18,19,22,23]. Research shows that low-SEP populations often have different views on health and the possible benefits of healthy behaviors compared to high-SEP populations [24-27]. Studies find that low-SEP populations have less time, more stress, and limited financial capacities to implement healthy behaviors [26]. Next to this, it is theorized that health beliefs often find origin in the health beliefs of previous generations [24-27]. In the context of low SEP, the expectation of a shorter life and the belief that their own behavior has limited influence on their longevity pose barriers to adhering to healthy behaviors [20]. Therefore, the digitalization of health care could seriously impact the access and use of health care for those who need it most [5,6,13,28-31].

National digital connectivity and policies that stimulate the transition toward a digital health care system could improve the implementation and accessibility of eHealth. The European Union and its member states deploy policies to realize the digital transition of health care systems [5,6]. Some states, such as the Netherlands, have been experimenting with digital health care for over a decade [4,6]. Most Dutch households (98%) have fast broadband coverage (2020) and 88% of the Dutch population uses mobile broadband (2019), which indicates the use of a mobile phone or other device with mobile internet access [5]. The level of connectivity in the Netherlands could facilitate the implementation of digital health care [5].

Several countries have developed national eHealth monitoring programs to monitor the uptake and effects of eHealth among health care professionals and citizens [32-34]. From a citizen's perspective, the monitoring programs focus on the use and evaluation of eHealth that involve citizen interaction [32-34]. This includes applications such as websites, apps, and wearables that citizens can use independently or involve eHealth tools that facilitate digital communication with a health care professional, such as video calls or messaging via patient portals [32-34]. This study is based on secondary data analysis of the Dutch eHealth monitoring program (2021) [29], which collects data about access, use, and perceived benefits from eHealth among Dutch citizens through questionnaires.

There is still limited understanding of the relationship between SEP and eHealth access, use, and perceived benefit and that understanding is generally limited to either access, use, or perceived benefit from eHealth, specific eHealth apps or specific subpopulations. Research showed the relation between SEP and the use of personal health records [35] and mobile apps [36]. Other research focuses on either the benefit from [37] or the use of eHealth [38,39] or specific patient groups such as cancer survivors [40,41], or citizens bound to specific locations [42]. To our knowledge, insight into differences in eHealth access, use, and perceived benefits and how different indicators for SEP display these differences within the Dutch general population aged 18 years and older are largely unknown. This study aims to assess differences in eHealth-related access, usage, and perceived benefits for different socioeconomic populations, based on education, standardized income, and socioeconomic status (SES) level of the neighborhood in the Netherlands. The findings of this study give insight into the disparities in access to, use of, and perceived benefit from eHealth in a highly connected country with an increasingly digitalized health care system. The results are insightful for other contexts that aim for or experience the same ambitions to transition to a digital health care system.

Methods

Panel

Data from the Dutch Health Care Consumer Panel (DHCCP) were used [43]. The DHCCP is a panel managed by Nivel (the Netherlands Institute for Health Services Research) and currently (as of September 2023) consists of approximately 11,500 panel members aged 18 years and older [43]. For this study, a study sample of 1.500 panel members was drawn by researchers from DHCCP. The study sample was representative of the Dutch population aged 18 years and older regarding age and sex [43]. Background characteristics of panel members, including their sex, age, level of education, net monthly income per household, and 4-digit postal codes were known. The panel was periodically renewed to ensure representative samples of the adult population in the Netherlands can be drawn. New panel members were recruited by buying an address file from an address supplier [43]. As a result, possible new members were sampled at random from the general population in the Netherlands [43]. The panel could only be joined through invitation. It was not possible for people to sign up on their own initiative [43]. Upon membership, panel members were informed of the purpose, scope, method,

and use of the panel [43]. Based on this information, participants could give permission to participate in the panel [43].

Ethical Considerations

According to the Dutch legislation, neither obtaining informed consent nor approval by a medical ethics committee was obligatory for doing research within the DHCCP [43,44]. Data analysis was conducted with pseudonymized data, according to the privacy regulations of the DHCCP, in compliance with the General Data Protection Regulation [44]. The privacy of the panel members was protected. All data were carefully stored by Nivel [43]. Personal information such as addresses was stored separately from the data of the questionnaires [43]. The privacy of the panel members in the study sample was guaranteed by DHCCP [43]. The researcher (LS) who analyzed the data had no access to the personal information of the panel members [43]. A written or digital informed consent was obtained at the time of registration of a new member to the panel [43]. Panel members were asked to participate approximately 4 or 5 times per year [43]. Participation was voluntary.

Data Collection

Data on the population's perspective on eHealth were collected via the DHCCP as part of a larger monitoring study into the perceptions, experiences, and usage of eHealth in the Netherlands [43]. A questionnaire was developed and reviewed by a team of representatives from the health care field in the Netherlands. The questionnaire was based on earlier distributed questionnaires of the monitoring study and was adjusted to reflect market developments [34]. The questionnaire was distributed via email and post (according to the preferences of the panel members) in May 2021. A digital reminder was sent after 1, 2, and 3 weeks after the start of the questionnaire and 1 written reminder was sent after 2 weeks. Panel members had 4 weeks to respond.

Socioeconomic Position Indicators

Overview

The concept of SEP is complex, as it results from the interaction between individual, social, economic, cultural, and societal factors [45,46]. In this study, 3 different operationalizations of SEP were used to study the digital divide in the context of eHealth—education as a historic starting point, standardized household income as a measure of current wealth, and SES level of the neighborhood to include environmental influence [46-49].

Education

The education levels were defined as low (none, primary school, or prevocational education); medium (secondary or vocational education level 1, 2, 3, or 4); and high (professional higher education or university) [50].

Standardized Income

Standardized income was defined as the net monthly income of the household adjusted for number of household members. The net income was converted to the equivalent of the net income of a single adult household by using equivalence factors from Statistics Netherlands (the Dutch Institute for Population Statistics) [51]. Some respondents acknowledged having children or other adults living in their household apart from their partner or children older than 18 years of age but did not specify the number. In the case of an unknown amount of children, 1.57 children were assumed. In case of an unknown amount of extra adults 1 extra adult was assumed. Assumptions were derived from the averages in Dutch households [50,52,53]. Information gathered from the panel members about their monthly net income was in ranges, and the mean of the range was taken as the monthly net income. Standardized income was divided into three categories (1) low (between €0 and €1659 [US $1718.56] per month; The used conversion rate was applicable on May 1, 2021), (2) medium (between €1660 [US $1719.59] and €2332 [US $2415.72] per month), and (3) high (more than €2332 [US $2415.72] per month). The categories were derived from the quartile distribution of the net income of the Dutch households (2020) [54].

Socioeconomic Status Level of the Neighborhood

The SES level of the neighborhood of all respondents was determined using the Social Economic Status-Wealth Education Employment (SES-WOA) score (2019) from the Statistics Netherlands. The SES-WOA score was based on the wealth, educational status, and recent employment history of households in the neighborhood [55,56]. The SES-WOA score was matched to the respondent by the 4-digit postal code. The neighborhoods of the respondents were categorized as (1) low (first tertile of SES score: –0.89 to 0.042), (2) medium (second tertile of SES score: 0.043-0.21), and (3) high (third tertile of SES score: 0.21-0.71). The average score of Dutch neighborhoods was 0.092 (SD 0.23; range –0.89 to 0.71) [55,56].

eHealth and the Digital Divide

For the interpretation of the data, the digital divide model was used. The digital divide model published by van Dijk et al [57] conceptualized that individuals' SEPs influence the available resources to access, use, and benefit from new digital media. In this study, the digital media in focus was eHealth.

Two concepts of eHealth were examined: (1) eHealth in general and (2) websites, apps, and wearables. Items in the questionnaire that informed these 2 concepts were matched to the 3 levels of the digital divide model [57], namely level 1—access, level 2—use, and level 3—perceived benefit. Operationalization of the variables measuring access, use, and perceived benefit for eHealth in general and websites, apps, and wearables can be found in Multimedia Appendix 1.

First, eHealth in general was studied to gain insight into an overall interest toward digital apps in health care. The levels of digital divide that were studied are access and the perceived benefit. Access was measured by motivation. Here respondents were asked what their general thoughts are about digital apps used in health care. The perceived benefit was operationalized by measuring to what extent the respondents perceived themselves as making more conscious decisions about their health as a result of eHealth use.

Second, eHealth in terms of websites, apps, and wearables was studied to gain insight into the use of these specific eHealth tools to improve health or provide support in coping with a disease. The digital divide levels that were taken into account

are access and use. Access was measured by motivation and physical access. Motivation was measured by asking the respondents if they have used or would like to use websites, apps, and wearables for their health. For physical access, the respondents were asked if they have access to an electronic device with internet.

The digital divide concept use was measured by barriers in use, diversity of use, and frequency of use. For the concept barriers in use, the respondents were asked if they experienced barriers in the use of websites, apps, and wearables. The use of websites, apps, and wearables was further operationalized in 2 variables—diversity of use and frequency of use. For these variables, the respondents were asked about 16 different websites, apps, or wearables if they used the app (once or more than once). Diversity entails the variety of websites, apps, and wearables used, while the frequency of use operationalizes the number of times (more than once) the apps were used.

Statistical Analysis

Descriptive analyses were used to describe the demographics and the outcome of the variables measuring the access, use, and perceived benefit for eHealth in general and websites, apps, and wearables. The variables measuring access, use, and perceived benefit were constructed by combining items from the original questionnaire. The operationalization of these variables can be found in Multimedia Appendix 1.

The outcomes were stratified by the 3 variables of SEP—educational level, standardized income, and SES level of the neighborhood. The differences in eHealth access and usage between SEP populations were investigated using logistic regression analysis. Ordered logistic regression was used for testing the differences in perceived benefit between SEP populations. Age was included in the analysis to test for confounding, as age is associated with both health and familiarity and use of digital media [10,58,59]. The correlation between the independent variables was determined via the Spearman rank order correlation coefficient. Data analysis was conducted using the Stata Statistical Software release (version 16.1; StataCorp) [60]. A $P<.05$ was considered statistically significant. Variables of physical access and diversity of use were not included in the logistic regression analysis because there were too few cases in outcome categories to meet the assumptions of the logistic regression analysis. The univariate outcomes are presented in Multimedia Appendix 2.

Results

Sample Characteristics

In total, 849 panel members responded to the questionnaire, resulting in a response rate of 56.6% (849/1500). Among the panel population of 1500, 8.9% (133/1500), 40% (600/1500), and 48.7% (731/1500) had a low, medium, and high level of education, respectively. Of these groups, 55.6% (74/133), 59.3% (356/600), and 54.3% (397/731) responded to the questionnaire. Regarding standardized income 35.8% (537/1500), 31.8% (477/1500), and 27.9 (419/1500) had low, medium, and high levels of standardized income, respectively. Of these groups, 57.2% (307/537), 56.4% (269/477), and 54.9% (230/269) had responded to the questionnaire. Finally, for SES-level of the neighborhood, 42.8% (642/1500), 32.6% (489/1500), and 23.1% (346/1500) had low, medium, and high levels of SES-level of the neighborhood in the panel population, respectively. Of these groups, 56.1% (360/642), 58.1% (284/489), and 55.2% (191/346) responded to the questionnaire.

The demographics of the study population can be found in Table 1. An overview of the study population stratified by education, standardized income, and SES level of the neighborhood can be found in Multimedia Appendix 3. Overall, the sample contained the same distribution of males or females as in the general population. When stratified for age category, our sample contained slightly more respondents aged 40 years and older and fewer respondents aged 18-39 years as compared to the general population. A frequency table of males and females in 3 age categories from the study population and the general Dutch population can be found in Multimedia Appendix 4 [61]. Compared to the Dutch general population, the study population had more males (11.3% general population, 14.9% study population) and females (13.1% general population, 15.4% study population) in the age category of 65 years and older. In terms of sex, the distribution was equal between both populations (Dutch general population of 49.3% male and 50.7% female, study population male 48.7% and female 51.3%). The mean age was 54 (SD 16.96) years. The lowest number of respondents was in the low educational level subpopulation (74/849, 8.72%) and the highest in the high educational level subpopulation (397/849, 46.76%). A high educational level was significantly and positively associated with a high standardized income level, as indicated by the correlations of educational level and standardized income level: $\rho=0.37$ ($P<.001$), educational level and SES level of the neighborhood: $\rho=0.021$ ($P=.55$), and standardized income level and SES level of the neighborhood: $\rho=0.035$ ($P=.33$) [62].

Table 1. Demographic description of the study population (n=849). Study population was sampled (2021) from a representative population (N=1500) of general Dutch population aged 18 years and older.

	Total (n=849)
Sex, n (%)	
Male	413 (48.6)
Female	435 (51.2)
Missing	1 (0)
Age (years)	
Mean (SD)	54 (16.96)
Range	19-92
Average household size	
Mean (SD)	2.29 (1.11)
Range	1-7
Number of households with children younger than 18 years, n (%)	
No	604 (71.1)
Yes	235 (27.7)
Missing	10 (1.2)
Number of households with children older than 18 years, n (%)	
No	763 (89.9)
Yes	76 (9)
Missing	10 (1.2)
Education	
Mean (SD)	2.39 (0.65)
Range	1-3
Low, n (%)	74 (8.7)
Medium, n (%)	356 (41.9)
High, n (%)	397 (46.8)
Missing, n (%)	22 (2.6)
Standardized income	
Mean (SD)	1.90 (0.81)
Range	1-3
Low, n (%)	307 (36.2)
Medium, n (%)	269 (31.7)
High, n (%)	230 (27.1)
Missing, n (%)	43 (5.1)
SES[b] level of the neighborhood	
Mean (SD)	1.80 (0.79)
Range	1-3
Low, n (%)	360 (42.4)
Medium, n (%)	284 (33.5)
High, n (%)	191 (22.5)
Missing, n (%)	14 (1.7)

[a]Education level—low (none, primary school or prevocational education); medium (secondary or vocational education level 1, 2, 3, or 4); and high (professional higher education or university). Standardized income was divided into 3 categories—low (between €0 and €1659 [US $1718.56] per month); medium (between €1660 [US $1719.59] and €2332 [US $2415.72] per month); and high (more than €2332 [US $2415.72] per month). The

SES level of the neighborhood was determined using the Social Economic Status-Wealth Education Employment (SES-WOA) score (2019) from Statistics Netherlands. The SES-WOA score was based on the wealth, educational status, and recent employment history of households in the neighborhood [55,56]. Categories: low (first tertile of SES score: –0.89 to 0.042); medium (second tertile of SES score: 0.043-0.21); and high (third tertile of SES score: 0.21-0.71). Not all values add up to 100% due to missing values.

[b]SES: socioeconomic status.

Results of the Relation Between the Digital Divide Levels and the Socioeconomic Position Indicators

Overview

An overview of the measured and analyzed variables can be found in Figure 1. The frequencies of these outcomes can be found in Multimedia Appendix 2. Table 2 presents a descriptive overview of access, use, and perceived benefit, stratified by SEP indicators. The results showed that the outcome differed the most when the population was stratified by education. Figure 2 presents the associations between access; use; perceived benefit; low, medium, and high education; standardized income; and SES level of the neighborhood populations. The underlying data used for Figure 2 is presented in Multimedia Appendix 5.

Figure 1. Overview of the concepts of access, use, and perceived benefit, the variables matched to these concepts and the socioeconomic position indicators used in this study. Legend indicates whether the variables were matched for eHealth in general (bold), websites apps and wearables (underlined) or both. Second, the legend indicates whether the variables were viable, either included (black) or not included (grey), for logistic (ordered) regression analysis [57]. SES: socioeconomic status.

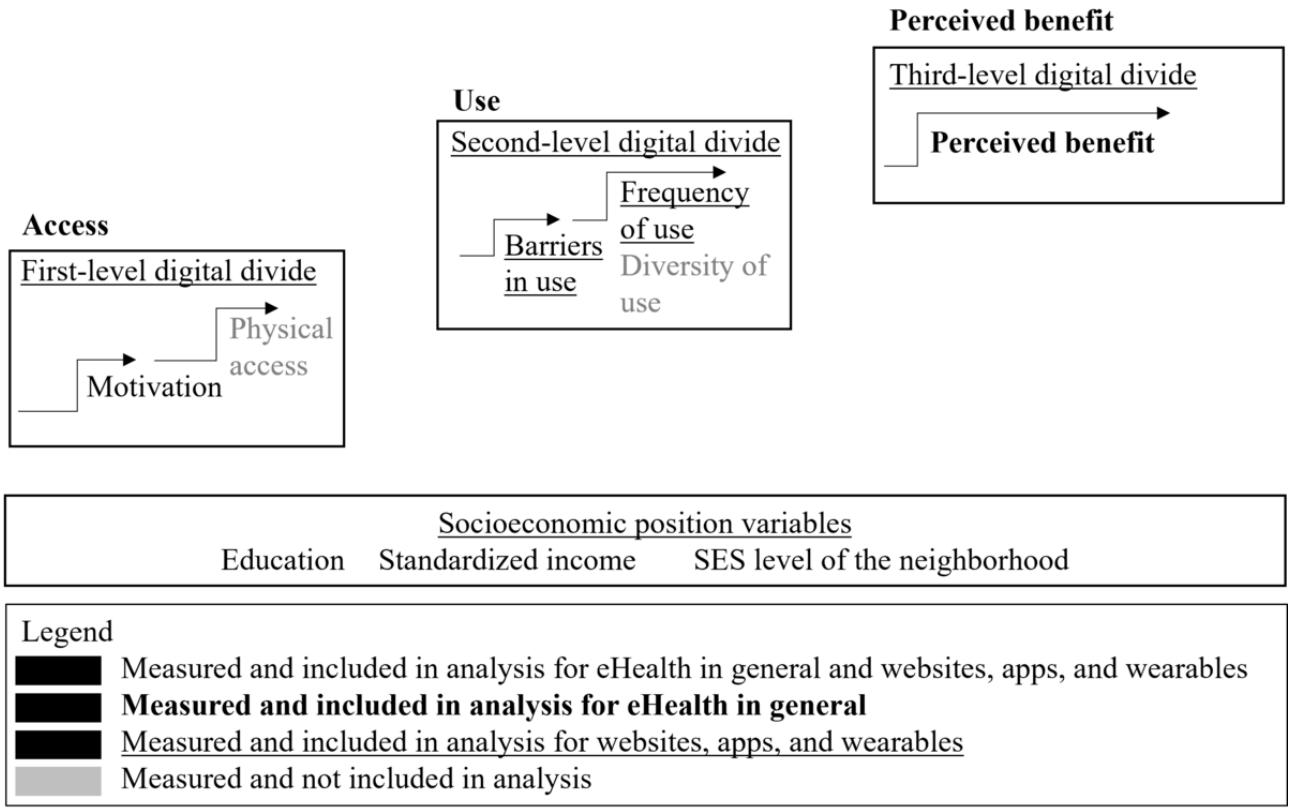


Figure 2. The relation between access, use, and benefit of eHealth and low, medium, and high levels of 3 SEP indicators—education, standardized income, and SES-level of the neighborhood. Results from a questionnaire (2021) answered by a representative study population (n=849) for the general Dutch population aged 18 years and older. Access: motivation, use: barriers in use and frequency of use, and perceived benefit: perceived benefit. For each SEP indicator, a bar graph of the results of logistic (ordered) regression analysis was presented. Each bar shows the odds ratio (OR) and the 95% CI for the difference between SEP levels for each digital divide concept. For each SEP indicator, the comparisons made were medium compared to low (dark grey), high compared to low (light grey), and high compared to medium (medium grey). The number of respondents (n) included in the analysis for each digital divide concept was indicated. The variable barriers in use were recoded to ensure that a positive score (1) reflected the outcome: no experienced barriers. Positive ORs (OR>1) should be interpreted that the primary group in the comparison, in the case of this study either medium or high SEP, had more likelihood of not experiencing barriers in use. *P>.05, **P>.01, ***P>.001. SEP: socioeconomic position; SES: socioeconomic status.

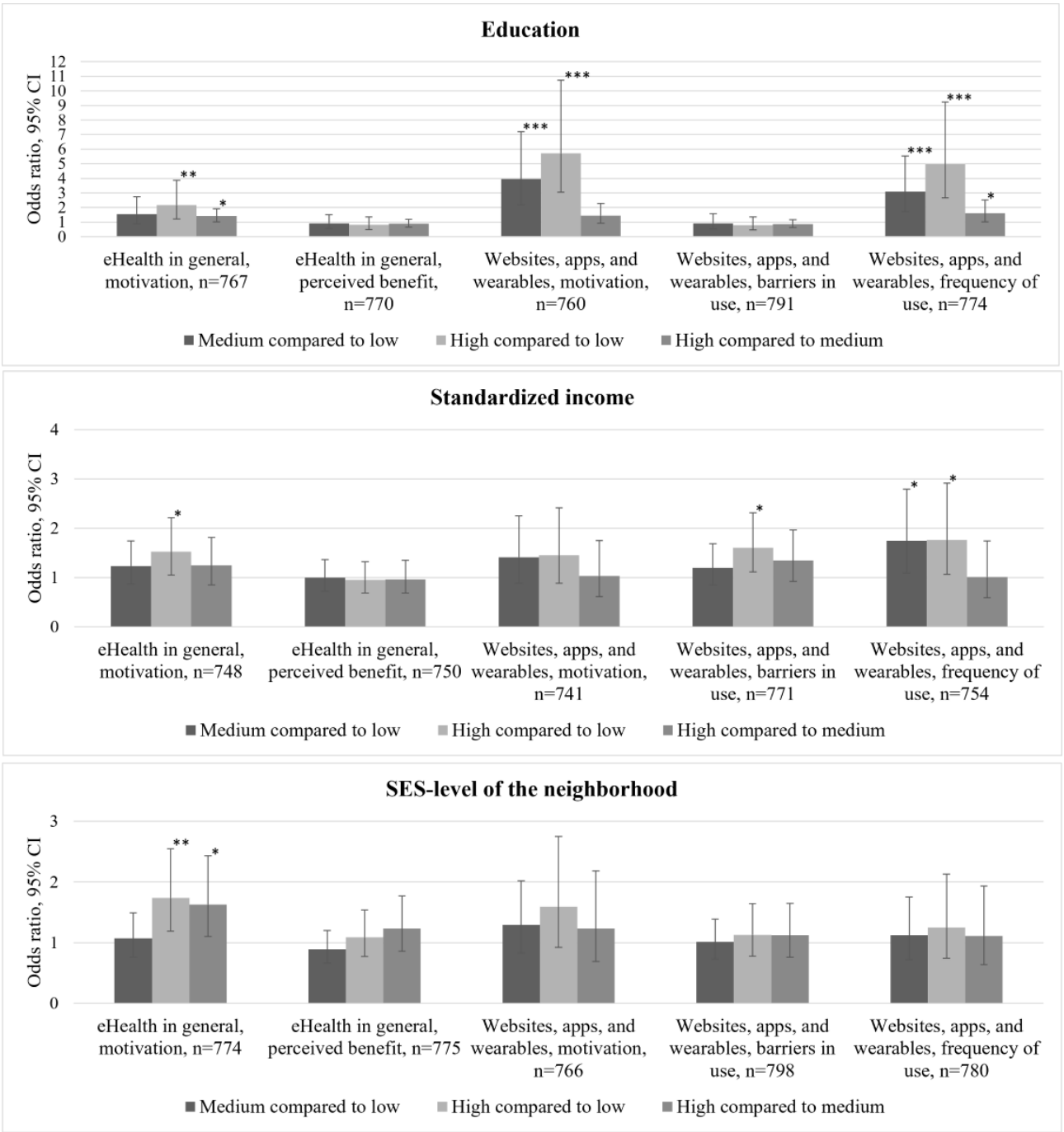


Table 2. Frequency distribution of access, use, and perceived benefit for (1) eHealth in general and (2) websites, apps, and wearables, stratified by education, standardized income, and SES[a] level of the neighborhood. Frequencies are derived from results of a questionnaire (2021) conducted among a sample of the general Dutch population (N=1500), final study population (n=849)[b,c].

	Total	Education			Standardized income			SES level of the neighborhood		
		Low	Med[d]	High	Low	Med	High	Low	Med	High
Population, n	849	74	356	397	307	269	230	360	284	191
eHealth in general										
Access—motivation, n (%)										
No motivation	386 (45.5)	46 (62.2)	177 (49.7)	152 (38.3)	153 (49.8)	124 (46.1)	90 (39.1)	174 (48.3)	141 (49.6)	68 (35.6)
Motivation	401 (47.2)	21 (28.4)	153 (43.0)	218 (54.9)	131 (42.7)	127 (47.2)	123 (53.5)	157 (43.6)	127 (44.7)	107 (56)
Perceived benefit—perceived benefit, n (%)										
Totally disagree—disagree	211 (24.9)	17 (23.0)	80 (22.4)	108 (27.2)	70 (22.8)	68 (25.3)	65 (28.3)	85 (23.6)	76 (26.7)	46 (24.1)
Not agree nor disagree	370 (43.6)	33 (44.5)	171 (48.0)	159 (40.1)	144 (46.9)	119 (44.2)	89 (38.7)	158 (43.9)	134 (47.2)	75 (39.3)
Agree—totally agree	205 (24.1)	17 (23.0)	81 (22.8)	104 (26.2)	69 (22.5)	65 (24.2)	61 (26.5)	87 (24.2)	61 (21.5)	53 (27.7)
Websites, apps, and wearables										
Access—motivation, n (%)										
No motivation	143 (16.8)	35 (47.3)	62 (17.4)	41 (10.3)	59 (19.2)	43 (16.0)	34 (14.8)	65 (18.1)	50 (17.6)	23 (12.0)
Motivation	636 (75.0)	29 (39.2)	265 (74.4)	328 (82.6)	221 (72.0)	206 (76.6)	178 (77.4)	263 (73.1)	215 (75.7)	150 (78.5)
Use—barriers in use, n (%)										
Experienced barriers	339 (40.0)	29 (39.2)	140 (39.3)	159 (40.1)	135 (44.0)	109 (40.5)	76 (33.0)	145 (40.3)	119 (41.9)	71 (37.2)
No experienced barriers	472 (55.6)	39 (52.7)	201 (56.5)	223 (56.2)	157 (51.1)	149 (55.4)	145 (63.0)	197 (54.7)	156 (54.9)	110 (57.6)
Use—frequency of use, n (%)										
No frequent use	131 (15.4)	29 (39.2)	60 (16.9)	38 (9.6)	59 (19.2)	36 (13.4)	29 (12.6)	57 (15.8)	47 (16.5)	24 (12.6)
Frequent use	662 (78.0)	34 (45.9)	274 (77.0)	339 (85.4)	222 (72.3)	219 (81.4)	189 (82.2)	277 (76.9)	226 (79.6)	149 (78.0)

[a]SES: socioeconomic status.

[b]The variables physical access and diversity of use for websites, apps, and wearables were not presented in Table 2 because these variables could not meet the assumptions for logistic regression analysis. The frequencies of these outcomes can be found in Multimedia Appendix 2.

[c]Not all values add up to 100% due to missing values.

[d]Med: medium level.

Access and Motivation

For eHealth in general, as well as for websites, apps and wearables, differences in motivation were found between different levels of education. Differences in motivation were most profound between low or medium versus highly educated respondents in both eHealth in general (odds ratio [OR] 2.18, 95% CI 1.22-3.88) and websites, apps, and wearables (OR 5.72, 95% CI 3.06-10.72). Regarding standardized income, a difference in motivation for eHealth in general was found between high and low standardized incomes (OR 1.52, 95% CI 1.05-2.21). A significant difference in motivation between low (OR 1.74, 95% CI 1.19-2.55) and medium (OR 1.63, 95% CI 1.1-2.43) versus high SES level of the neighborhood was found for eHealth in general.

Use, Barriers in Use, and Experienced Barriers

High standardized income was associated with no experience of barriers in use in comparison to low and medium levels of standardized income. This implies that fewer respondents with a high standardized income experienced barriers while using eHealth websites, apps, and wearables (OR 1.60, 95% CI 1.11-2.31). The frequency of eHealth use also differed between respondents with a low, medium, or high level of education,

with the most significant difference between high and low educational levels (OR 4.96, 95% CI 2.66-9.24). In terms of standardized income, there were significant differences between low and high (OR 1.76, 95% CI 1.06-2.91) and low and medium (OR 1.74, 95% CI 1.09-2.79) standardized income levels. High SEP respondents were more likely to frequently use an eHealth app, website, or wearable compared to medium or low SEP respondents.

Perceived Benefit

There were no significant differences found regarding the perceived benefits between low, medium, or high-SEP populations.

Discussion

Principal Findings

This study shows that low SEP respondents have less access to and use of eHealth compared to medium or high SEP respondents. The most significant digital divide observed in this study is related to educational background. The results of this paper contribute on a population level to previous findings that the access and use of eHealth has a socioeconomic gradient. Additionally, the results emphasize that SEP indicators cannot be used interchangeably to assess eHealth access and use.

The results of this study highlight that, across all 3 SEP indicators, the most substantial differences are found in access through motivation. Respondents from higher socioeconomic categories expressed greater motivation to use eHealth, including websites, apps, and wearables, in comparison to those from lower SEPs. Health equity researchers emphasize that comprehending how people perceive health and health care is a complex issue influenced by several societal, contextual, social, and biological individual factors. Therefore, a multidimensional and multicausal approach is necessary to comprehend these disparities [24,63,64]. Weiss et al [28] discuss existing literature and multiple theories as to why differences in eHealth access and use exist between socioeconomic groups. The literature described that the social position of individuals and the influence of the context and organizations surrounding the individual play a role in whether individuals choose to consume digital health care or not [28]. Other literature emphasized that the diffusion of digital health care in society will decrease the digital divide gap as the low-SEP population is assumed to be the latest to adopt [28]. However, the role of health care organizations, social policies, and political decisions that impact individuals' motivations toward eHealth has not been adequately studied. Current national Dutch policies concerning eHealth are focused on the development and interoperability of eHealth apps, the digital skills of health care professionals and the use of eHealth by older people at home, resulting in eHealth becoming an essential part of the health care system [65]. In countries where health care digitalization is progressing, it would be valuable to examine the potential influence of the government, health care organizations, and businesses on the motivation of low socioeconomic populations toward eHealth. Such research could offer valuable insights into the societal and policy changes required to make eHealth more appealing to low-SEP populations.

In this study, use was examined by studying the barriers in use and frequency of use of eHealth websites, apps, and wearables. The results demonstrated that respondents with a high standardized income level infrequently experienced barriers to using eHealth websites, apps, and wearables compared to lower standardized income levels. Previous studies show that highly educated people often have higher health literacy levels and digital skills and are more in contact with the digital world via their education or profession [10,31,59,66-70]. Other studies show that in the development of eHealth new apps are often pilot-tested by highly educated respondents and, therefore, might be more tailored to the needs of highly educated individuals [9,71,72]. In contrast, other studies point out the variety of health behaviors and health care use within SEP groups. De Boer et al [73] show that low SEP groups have more health care costs but that healthy lifestyle behavior such as smoking or being member of a sports club are attributed greatly to the variety in health care use in each socioeconomic group. In the light of eHealth, Agachi et al [74] showed that the user interface and the type of eHealth offered attributes to the use of eHealth between socioeconomic groups. Results revealed that for the same eHealth program, people living in a low socioeconomic neighborhood use the app-based tool more than people living high in a socioeconomic neighborhood. For the web-based version, results show the opposite emphasizing the importance of the user interface and the accessibility to digital devices, as is also theorized in the digital divide model [49,74]. Both De Boer [73] and Agachi [74] show that behavior-related and technical-related factors play a role in the use of health care and eHealth. The results of this study and other studies showed that understanding and creating insight into the existence of and possible solutions for health disparities is dependent on multiple dimensions. Future research into how the socioeconomic gradient in eHealth access and use are associated with other behavioral and technical factors is important to create an in-depth understanding of disparities in eHealth access and use that can inspire research, policy, and practice.

Results pointed out that high education and high standardized income levels were associated with frequent use of websites, apps, and wearables. This is in accordance with previous research, which shows that a high level of education and income is found to be associated with more access to and use of eHealth [12,13,31]. Surprisingly, no difference was found in the perceived benefits (making more conscious decisions in health because of eHealth use) in any of the SEP indicators, although the frequency of use of websites, apps, and wearables was high (78%) and did show significant differences. This implies that respondents with more frequent use of websites, apps, and wearables had the same perceived benefit, namely, making more conscious decisions due to eHealth use, compared to respondents who have not used websites, apps, and wearables once. Evaluation studies show that high SEP respondents have better outcomes from eHealth use than low SEP respondents [10,30,31,59]. The results of this study might indicate that in the real-life context, even though current eHealth apps might be more suitable for highly educated individuals, eHealth is not used appropriately or with similar discipline as in the clinical trial context.

Strengths and Limitations

The strengths of this study were the use of a large and representative sample of the Dutch population and the use of 3 SEP operationalizations to provide a broad insight into the effects of SEP on the digital divide. This study also has some limitations. Despite a large number of respondents, the skewed distribution of outcomes across SEP levels hampered the performance of multivariable logistic regression analysis [62]. Next to this, the SEP indicators used are focused on social demographic and economic aspects of SEP. Cultural and other social aspects, such as social networks and cultural background, have not been taken into account.

The questionnaire used in the study was not designed to measure all the different aspects of the digital divide model and is a secondary analysis of the data gathered. Although the majority of digital divide levels could be well matched with questionnaire items, data on the second level digital divide for eHealth in general and the third digital divide level for websites, apps, and wearables was lacking. In some cases, with emphasis on the concept barriers in use for websites, apps, and wearables more in-depth insights into the digital divide levels would have been desirable to improve the validation of the findings. Van Dijk [57] provides concepts to further define the digital divide levels. Access, the first level of digital divide, is described to entail the concepts of motivational access and physical access. Use, the second level of digital divide entails the concepts of digital skills and usage. Usage here encompasses both frequency and diversity of digital media use. Perceived benefit, the third level digital divide, is conceptualized by personal outcomes that are a result of the use of digital media [57]. In this study, the concept barriers in use were used instead of digital skills, as experienced barriers are not limited to barriers formed by a lack of digital skills.

Additionally, for the first and the second level digital divide, the technical design and the information and communication technology of eHealth are of importance [57]. The technical design and the information and communication technology imply factors such as accessibility, usability, mobility, quality, and accessibility of internet access and automation (self-learning devices or software tailored to serve the consumer better and automatically) of devices and apps. These factors are important to facilitate adherence and appropriate use of eHealth apps [57]. The questionnaire provided no insight into these factors.

Conclusions

The results of this study revealed that differences in motivation for eHealth use are most profound between different socioeconomic populations in the Dutch society, in which low-educated people are likely to be disadvantaged. A successful transition toward digital health care is a social issue that is dependent on the motivation to use eHealth and specific apps. It is imperative that future studies within academia and within the health care field focus on the motivations and needs associated with digital health care, specifically for low-SEP populations. Research on the societal changes stemming from the digital health care transition and the technical and design studies of digital care apps in single-intervention studies are both vital for the realization of an inclusive and comprehensive digital health care system. If eHealth takes a predominant role in the Dutch health care system, it might affect access and use of health care for the citizens who need it the most. The results of this study underscore the importance of policies aimed at facilitating and supporting low-SEP populations in the use of eHealth to reduce differences in health.

Acknowledgments

The authors would like to thank E de Vries and L van Deursen for providing comments and suggestions. This study was funded by the Dutch Ministry of Health, Welfare, and Sport. The funders had no role in the study design, data collection and analysis, data interpretation, writing of the paper, or approval for publication.

Data Availability

The datasets generated and analyzed during this study are available upon reasonable request from Prof Judith D de Jong (j.dejong@nivel.nl), project leader of the Dutch Health Care Consumer Panel, or the panel's secretary (conusmentenpanel@nivel.nl).

Conflicts of Interest

None to declare.

Multimedia Appendix 1
Operationalization of access, use, and perceived benefit of eHealth in general and websites, apps, and wearables.
[DOCX File , 24 KB - formative_v9i1e49585_app1.docx]

Multimedia Appendix 2
Frequency table of the outcomes of access, use, and perceived benefit for eHealth in general and websites, apps, and wearables gathered via a questionnaire (2021).
[DOCX File , 24 KB - formative_v9i1e49585_app2.docx]

Multimedia Appendix 3
Demographic description of the study population and the study population stratified by levels of education, standardized income, and SES level of the neighborhood.
[DOCX File , 31 KB - formative_v9i1e49585_app3.docx]

Multimedia Appendix 4
Frequency table of males and females aged above 18 in three age categories of the Dutch general population (2021) and the study population (2021).
[DOCX File , 14 KB - formative_v9i1e49585_app4.docx]

Multimedia Appendix 5
Results from logistic (ordered) regression analysis.
[DOCX File , 21 KB - formative_v9i1e49585_app5.docx]

References

1. Rutten P, Pruim J, Zijl NV, Merckelbach S. Digitale zorg in Nederland. Digitale zorg in Nederland. 2020. URL: https://www.mckinsey.com/nl/our-insights/digitale-zorg-in-nederland [accessed 2023-05-05]
2. Abernethy A, Adams L, Barrett M, Bechtel C, Brennan P, Butte A, et al. The promise of digital health: then, now, and the future. NAM Perspect 2022 Jun 27:202206e [FREE Full text] [doi: 10.31478/202206e] [Medline: 36177208]
3. Pagliari C, Sloan D, Gregor P, Sullivan F, Detmer D, Kahan JP, et al. What is eHealth (4): a scoping exercise to map the field. J Med Internet Res 2005 Mar 31;7(1):e9 [FREE Full text] [doi: 10.2196/jmir.7.1.e9] [Medline: 15829481]
4. Dutch digitisation strategy 2.0. Ministery of Economic Affairs and Climate Policy. 2019. URL: https://www.nederlanddigitaal.nl/documenten/publicaties/2019/11/13/english-version-of-the-dutch-digitalisation-strategy-2.0 [accessed 2023-05-05]
5. Digital economy and society index (DESI) 2021: The Netherlands. European Commission. 2021. URL: https://ec.europa.eu/newsroom/dae/redirection/document/80493 [accessed 2022-02-01]
6. 2030 digital compass: the European way for the digital decade. European Commission. 2021. URL: https://eufordigital.eu/library/2030-digital-compass-the-european-way-for-the-digital-decade/ [accessed 2022-02-01]
7. Boers SN, Jongsma KR, Lucivero F, Aardoom J, Büchner FL, de Vries M, et al. SERIES: eHealth in primary care. Part 2: Exploring the ethical implications of its application in primary care practice. Eur J Gen Pract 2019 Oct 30;26(1):26-32 [FREE Full text] [doi: 10.1080/13814788.2019.1678958] [Medline: 31663394]
8. European C. Protecting fundamental rights in the digital age - 2021 annual report on the application of the EU charter of fundamental rights. European Commission. 2021. URL: https://ec.europa.eu/commission/presscorner/detail/en/ip_21_6699 [accessed 2023-04-06]
9. Goedhart NS, Zuiderent-Jerak T, Woudstra J, Broerse JEW, Betten AW, Dedding C. Persistent inequitable design and implementation of patient portals for users at the margins. J Am Med Inform Assoc 2021 Feb 15;28(2):276-283 [FREE Full text] [doi: 10.1093/jamia/ocaa273] [Medline: 33463691]
10. Latulippe K, Hamel C, Giroux D. Social health inequalities and eHealth: a literature review with qualitative synthesis of theoretical and empirical studies. J Med Internet Res 2017 Apr 27;19(4):e136 [FREE Full text] [doi: 10.2196/jmir.6731] [Medline: 28450271]
11. McAuley A. Digital health interventions: widening access or widening inequalities? Public Health 2014 Dec;128(12):1118-1120. [doi: 10.1016/j.puhe.2014.10.008] [Medline: 25458115]
12. Kontos E, Blake KD, Chou WYS, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. J Med Internet Res 2014 Jul 16;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]
13. Perez SL, Kravitz RL, Bell RA, Chan MS, Paterniti DA. Characterizing internet health information seeking strategies by socioeconomic status: a mixed methods approach. BMC Med Inform Decis Mak 2016 Aug 09;16:107. [doi: 10.1186/s12911-016-0344-x] [Medline: 27506607]
14. Scherer R, Siddiq F. The relation between students' socioeconomic status and ICT literacy: findings from a meta-analysis. Comput Educ 2019;138:13-32. [doi: 10.1016/j.compedu.2019.04.011]
15. Azzopardi-Muscat N, Sørensen K. Towards an equitable digital public health era: promoting equity through a health literacy perspective. Eur J Public Health 2019;29(Supplement_3):13-17 [FREE Full text] [doi: 10.1093/eurpub/ckz166] [Medline: 31738443]
16. Wong BLH, Maaß L, Vodden A, van Kessel R, Sorbello S, Buttigieg S, European Public Health Association (EUPHA) Digital Health Section. The dawn of digital public health in Europe: implications for public health policy and practice. Lancet Reg Health Eur 2022;14:100316 [FREE Full text] [doi: 10.1016/j.lanepe.2022.100316] [Medline: 35132399]
17. Raad VVES. Een eerlijke kans op gezond leven. Raad voor Volksgezondheid en Samenleving. 2021. URL: https://www.raadrvs.nl/documenten/publicaties/2021/04/07/een-eerlijke-kans-op-gezond-leven [accessed 2021-10-13]

18. de Boer WIJ, Buskens E, Koning RH, Mierau JO. Neighborhood socioeconomic status and health care costs: a population-wide study in the Netherlands. Am J Public Health 2019;109(6):927-933. [doi: 10.2105/AJPH.2019.305035] [Medline: 30998412]
19. Loef B, Meulman I, Herber GM, Kommer GJ, Koopmanschap MA, Kunst AE, et al. Socioeconomic differences in healthcare expenditure and utilization in the Netherlands. BMC Health Serv Res 2021;21(1):643 [FREE Full text] [doi: 10.1186/s12913-021-06694-9] [Medline: 34217287]
20. Petrovic D, de Mestral C, Bochud M, Bartley M, Kivimäki M, Vineis P, et al. The contribution of health behaviors to socioeconomic inequalities in health: a systematic review. Prev Med 2018;113:15-31 [FREE Full text] [doi: 10.1016/j.ypmed.2018.05.003] [Medline: 29752959]
21. Phelan JC, Link BG, Tehranifar P. Social conditions as fundamental causes of health inequalities: theory, evidence, and policy implications. J Health Soc Behav 2010;51 Suppl:S28-S40. [doi: 10.1177/0022146510383498] [Medline: 20943581]
22. Agerholm J, Bruce D, Ponce de Leon A, Burström B. Socioeconomic differences in healthcare utilization, with and without adjustment for need: an example from Stockholm, Sweden. Scand J Public Health 2013;41(3):318-325. [doi: 10.1177/1403494812473205] [Medline: 23406653]
23. Lemstra M, Mackenbach J, Neudorf C, Nannapaneni U. High health care utilization and costs associated with lower socio-economic status: results from a linked dataset. Can J Public Health 2009;100(3):180-183 [FREE Full text] [doi: 10.1007/BF03405536] [Medline: 19507718]
24. Mackenbach JP. The persistence of health inequalities in modern welfare states: the explanation of a paradox. Soc Sci Med 2012;75(4):761-769. [doi: 10.1016/j.socscimed.2012.02.031] [Medline: 22475407]
25. Oude Groeniger J, Kamphuis CBM, Mackenbach JP, Beenackers MA, van Lenthe FJ. Are socio-economic inequalities in diet and physical activity a matter of social distinction? A cross-sectional study. Int J Public Health 2019;64(7):1037-1047 [FREE Full text] [doi: 10.1007/s00038-019-01268-3] [Medline: 31187165]
26. Pampel FC, Krueger PM, Denney JT. Socioeconomic disparities in health behaviors. Annu Rev Sociol 2010;36:349-370. [doi: 10.1146/annurev.soc.012809.102529] [Medline: 21909182]
27. Singh-Manoux A, Marmot M. Role of socialization in explaining social inequalities in health. Soc Sci Med 2005;60(9):2129-2133 [FREE Full text] [doi: 10.1016/j.socscimed.2004.08.070] [Medline: 15743660]
28. Weiss D, Rydland HT, Øversveen E, Jensen MR, Solhaug S, Krokstad S. Innovative technologies and social inequalities in health: a scoping review of the literature. PLoS One 2018;13(4):e0195447 [FREE Full text] [doi: 10.1371/journal.pone.0195447] [Medline: 29614114]
29. van der VR, van DL, Standaar LMB, Wouters EJM, Suijkerbuijk AWM, van TL. E-healthmonitor 2021: stand van zaken digitale zorg. 2022. URL: https://www.rivm.nl/documenten/e-healthmonitor-2021-stand-van-zaken-digitale-zorg [accessed 2023-05-05]
30. 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;18(1):148 [FREE Full text] [doi: 10.1186/s12966-021-01218-4] [Medline: 34753490]
31. Hansen AH, Bradway M, Broz J, Claudi T, Henriksen Ø, Wangberg SC, et al. Inequalities in the use of eHealth between socioeconomic groups among patients with type 1 and type 2 diabetes: cross-sectional study. J Med Internet Res 2019;21(5):e13615 [FREE Full text] [doi: 10.2196/13615] [Medline: 31144669]
32. Verhellen A, Jewell C, Van Gils M, Jacobs A, Steenberghs E. Ehealthmonitor Citizens. 2019. URL: https://researchportal.vub.be/en/publications/ehealthmonitor-2019-citizens [accessed 2023-09-13]
33. Vehko T, Ruotsalainen R, Hyppönen H, Aalto A, Terveyden ja hyvinvoinnin laitos. E-health and e-welfare of Finland: check point 2018. 2019. URL: https://oula.finna.fi/Record/oy.9917519693906252?lng=en-gb [accessed 2023-09-13]
34. Huygens MWJ, Voogdt-Pruis HR, Wouters M, Meurs MM, van Lettow B, Kleijweg C, et al. The uptake and use of telemonitoring in chronic care between 2014 and 2019: nationwide survey among patients and health care professionals in the Netherlands. J Med Internet Res 2021;23(5):e24908 [FREE Full text] [doi: 10.2196/24908] [Medline: 33938808]
35. Paccoud I, Baumann M, Le Bihan E, Pétré B, Breinbauer M, Böhme P, et al. Socioeconomic and behavioural factors associated with access to and use of personal health records. BMC Med Inform Decis Mak 2021;21(1):18 [FREE Full text] [doi: 10.1186/s12911-020-01383-9] [Medline: 33435970]
36. Bol N, Helberger N, Weert JCM. Differences in mobile health app use: a source of new digital inequalities? Inf Soc 2018;34(3):183-193. [doi: 10.1080/01972243.2018.1438550]
37. Van Deursen AJM, Helsper EJ. The third-level digital divide: who benefits most from being online. In: Robinson, L, Cotten SR, Schulz J, Hale TM, Williams A, editors. Communication and Information Technologies Annual. Leeds, England: Emerald Group Publishing Limited; 2015:29-52.
38. Chagpar AB. Sociodemographic factors affecting telemedicine access: a population-based analysis. Surgery 2022;171(3):793-798 [FREE Full text] [doi: 10.1016/j.surg.2021.08.059] [Medline: 34848076]
39. Aldhahir AM, Alqahtani JS, Althobiani MA, Alghamdi SM, Alanazi AF, Alnaim N, et al. Current knowledge, satisfaction, and use of E-Health mobile application (Seha) among the general population of Saudi Arabia: a cross-sectional study. J Multidiscip Healthc 2022;15:667-678 [FREE Full text] [doi: 10.2147/JMDH.S355093] [Medline: 35399807]

40. Moon Z, Zuchowski M, Moss-Morris R, Hunter MS, Norton S, Hughes LD. Disparities in access to mobile devices and e-health literacy among breast cancer survivors. Support Care Cancer 2022;30(1):117-126 [FREE Full text] [doi: 10.1007/s00520-021-06407-2] [Medline: 34236506]
41. Melhem SJ, Nabhani-Gebara S, Kayyali R. Digital trends, digital literacy, and E-Health engagement predictors of breast and colorectal cancer survivors: a population-based cross-sectional survey. Int J Environ Res Public Health 2023;20(2):1472 [FREE Full text] [doi: 10.3390/ijerph20021472] [Medline: 36674237]
42. Özdemir R, Bektemur G, Keles E, Baydili KN. Internet use, e-Heath literacy, and associated factors in Istanbul, Turkey: a cross-sectional study. J Consumer Health Internet 2023;27(1):1-11 [FREE Full text] [doi: 10.1080/15398285.2022.2129178]
43. Brabers AEM, de JJD. Nivel Consumentenpanel Gezondheidszorg Basisrapport met informatie over het panel 2022. NIVEL. 2022. URL: https://www.nivel.nl/nl/publicatie/nivel-consumentenpanel-gezondheidszorg-basisrapport-met-informatie-over-het-panel-2022 [accessed 2023-04-06]
44. Your research: is it subject to the WMO or not? Central Committee on Research Involving Human Subjects. URL: https://english.ccmo.nl/investigators/legal-framework-for-medical-scientific-research/your-research-is-it-subject-to-the-wmo-or-not [accessed 2024-06-13]
45. Giddens A. The Constitution of Society: Outline of the Theory of Structuration. Berkeley: University of California Press; 1984.
46. Bourdieu P. Outline of a Theory of Practice. Cambridge: Cambridge University Press; 1977.
47. Krieger N, Williams DR, Moss NE. Measuring social class in US public health research: concepts, methodologies, and guidelines. Annu Rev Public Health 1997;18:341-378. [doi: 10.1146/annurev.publhealth.18.1.341] [Medline: 9143723]
48. Scheerder A, van Deursen A, van Dijk J. Determinants of internet skills, uses and outcomes. A systematic review of the second- and third-level digital divide. Telematics Inform 2017;34(8):1607-1624. [doi: 10.1016/j.tele.2017.07.007]
49. Vrooman J, Boelhouwer J, Iedema J, van der T. Summary contemporary inequality. The Netherlands Institute for Social Research. 2023. URL: https://english.scp.nl/publications/publications/2023/03/07/summary-contemporary-inequality [accessed 2023-05-09]
50. Opleidingsniveau. Statistics Netherlands. 2019. URL: https://www.cbs.nl/nl-nl/nieuws/2019/33/verschil-levensverwachting-hoog-en-laagopgeleid-groeit/opleidingsniveau [accessed 2022-01-27]
51. What is my spendable income? Statistics Netherlands. 2018. URL: https://www.cbs.nl/en-gb/background/2008/50/what-is-my-spendable-income- [accessed 2023-04-06]
52. Geboorte; kerncijfers. StatLine. 2022. URL: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/37422ned/table [accessed 2022-02-20]
53. Ruim 140 duizend samenwonende volwassen kinderen met ouders op leeftijd. Centraal Bureau van de Statistiek. 2014. URL: https://www.cbs.nl/nl-nl/achtergrond/2014/27/ruim-140-duizend-samenwonende-volwassen-kinderen-met-ouders-op-leeftijd [accessed 2022-08-02]
54. Income distribution (standardised income). Centraal Bureau van de Statistiek. 2019. URL: https://www.cbs.nl/en-gb/visualisations/income-distribution [accessed 2021-08-04]
55. Statusscore per wijk en buurt o.b.v. welvaart, opleidingsniveau en arbeid. Centraal Bureau van de Statistiek. 2022. URL: https://www.cbs.nl/nl-nl/achtergrond/2022/14/statusscore-per-wijk-en-buurt-o-b-v-welvaart-opleidingsniveau-en-arbeid [accessed 2022-08-01]
56. Sociaal-economische status; scores per wijk en buurt, regio-indeling 2021. Centraal Bureau van de Statistiek. 2022. URL: https://www.cbs.nl/nl-nl/cijfers/detail/85163NED [accessed 2022-08-01]
57. Van Dijk J. The Digital Divide. Cambridge: Polity Press; 2020.
58. Elena-Bucea A, Cruz-Jesus F, Oliveira T, Coelho PS. Assessing the role of age, education, gender and income on the digital divide: evidence for the European union. Inf Syst Front 2020;23(4):1007-1021. [doi: 10.1007/s10796-020-10012-9]
59. Reiners F, Sturm J, Bouw LJW, Wouters EJM. Sociodemographic factors influencing the use of eHealth in people with chronic diseases. Int J Environ Res Public Health 2019;16(4):645 [FREE Full text] [doi: 10.3390/ijerph16040645] [Medline: 30795623]
60. Stata statistical software: release 16. StataCorp. 2019. URL: https://www.stata-press.com/data/r16/ [accessed 2024-06-20]
61. Bevolking op 1 januari en gemiddeld; geslacht, leeftijd en regio. Statistics Netherlands. 2021. URL: https://www.cbs.nl/nl-nl/cijfers/detail/03759ned?dl=39E0B [accessed 2024-05-15]
62. Akoglu H. User's guide to correlation coefficients. Turk J Emerg Med 2018;18(3):91-93 [FREE Full text] [doi: 10.1016/j.tjem.2018.08.001] [Medline: 30191186]
63. Øversveen E, Rydland HT, Bambra C, Eikemo TA. Rethinking the relationship between socio-economic status and health: making the case for sociological theory in health inequality research. Scand J Public Health 2017;45(2):103-112. [doi: 10.1177/1403494816686711] [Medline: 28078944]
64. Eikemo TA, Øversveen E. Social inequalities in health: challenges, knowledge gaps, key debates and the need for new data. Scand J Public Health 2019;47(6):593-597. [doi: 10.1177/1403494819866416] [Medline: 31512565]
65. Stimuleren gebruik e-health. Ministry of Public Health, Wellbeing and Sports. URL: https://www.rijksoverheid.nl/onderwerpen/e-health/overheid-stimuleert-e-health [accessed 2023-09-08]

66. Hiltz SR, Turoff M. Education goes digital: the evolution of online learning and the revolution in higher education. Commun ACM 2005;48(10):59-64. [doi: 10.1145/1089107.1089139]
67. Grossman LV, Masterson Creber RM, Benda NC, Wright D, Vawdrey DK, Ancker JS. Interventions to increase patient portal use in vulnerable populations: a systematic review. J Am Med Inform Assoc 2019;26(8-9):855-870 [FREE Full text] [doi: 10.1093/jamia/ocz023] [Medline: 30958532]
68. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
69. Bejaković P, Mrnjavac Z. The importance of digital literacy on the labour market. ER 2020;42(4):921-932. [doi: 10.1108/er-07-2019-0274]
70. Vasilescu MD, Serban AC, Dimian GC, Aceleanu MI, Picatoste X. Digital divide, skills and perceptions on digitalisation in the European union-towards a smart labour market. PLoS One 2020;15(4):e0232032 [FREE Full text] [doi: 10.1371/journal.pone.0232032] [Medline: 32324786]
71. Wagenaar KP, Hakim N, Broekhuizen BDL, Jaarsma T, Rutten FH, Hoes AW. Representativeness of participants in heart failure E-health trials: a report from the E-vita HF study. J Card Fail 2017;23(1):88-89 [FREE Full text] [doi: 10.1016/j.cardfail.2016.08.006] [Medline: 27565046]
72. Al-Dhahir I, Reijnders T, Faber JS, van den Berg-Emons RJ, Janssen VR, Kraaijenhagen RA, et al. The barriers and facilitators of eHealth-based lifestyle intervention programs for people with a low socioeconomic status: scoping review. J Med Internet Res 2022;24(8):e34229 [FREE Full text] [doi: 10.2196/34229] [Medline: 36001380]
73. de Boer WIJ, Dekker LH, Koning RH, Navis GJ, Mierau JO. How are lifestyle factors associated with socioeconomic differences in health care costs? Evidence from full population data in the Netherlands. Prev Med 2020;130:105929 [FREE Full text] [doi: 10.1016/j.ypmed.2019.105929] [Medline: 31778685]
74. Agachi E, Bijmolt THA, Mierau JO, van Ittersum K. Adoption of the website and mobile app of a preventive health program across neighborhoods with different socioeconomic conditions in the Netherlands: longitudinal study. JMIR Hum Factors 2022;9(1):e32112 [FREE Full text] [doi: 10.2196/32112] [Medline: 35107433]

Abbreviations

DHCCP: Dutch Health Care Consumer Panel
OR: odds ratio
SEP: socioeconomic position
SES-WOA: Social Economic Status-Wealth Education Employment
SES: socioeconomic status

Edited by A Mavragani; submitted 03.06.23; peer-reviewed by J Mierau, H Traino, PhD, MPH, M Earl; comments to author 29.08.23; revised version received 02.10.23; accepted 21.07.24; published 07.01.25.

Please cite as:
Standaar L, van Tuyl L, Suijkerbuijk A, Brabers A, Friele R
Differences in eHealth Access, Use, and Perceived Benefit Between Different Socioeconomic Groups in the Dutch Context: Secondary Cross-Sectional Study
JMIR Form Res 2025;9:e49585
URL: https://formative.jmir.org/2025/1/e49585
doi:10.2196/49585
PMID:

Original Paper

The Trifecta of Industry, Academic, and Health System Partnership to Improve Mental Health Care Through Smartphone-Based Remote Patient Monitoring: Development and Usability Study

C Neill Epperson[1], MD; Rachel Davis[1], MD; Allison Dempsey[1], PhD; Heinrich C Haller[1], BS; David J Kupfer[2], MD; Tiffany Love[1], PhD; Pamela M Villarreal[3], MS; Mark Matthews[4], PhD; Susan L Moore[5], PhD; Kimberly Muller[5], MSc, JD; Christopher D Schneck[1], MD; Jessica L Scott[1], MA, MPH; Richard D Zane[3,6], MD; Ellen Frank[2,4], PhD

[1]Department of Psychiatry, School of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, United States

[2]Department of Psychiatry, School of Medicine, University of Pittsburgh, Pittsburgh, PA, United States

[3]UCHealth CARE Innovations Center, Anschutz Medical Campus, University of Colorado Health System, Aurora, CO, United States

[4]Health Rhythms, New York, NY, United States

[5]CU Innovations, University of Colorado Anschutz Medical Campus, Aurora, CO, United States

[6]Department of Emergency Medicine, School of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, United States

Corresponding Author:
C Neill Epperson, MD
Department of Psychiatry
School of Medicine
University of Colorado Anschutz Medical Campus
1890 N Revere Ct
Suite 4003, Mail Stop F546
Aurora, CO, 80045
United States
Phone: 1 303 724 4940
Email: neill.epperson@cuanschutz.edu

Abstract

Background: Mental health treatment is hindered by the limited number of mental health care providers and the infrequency of care. Digital mental health technology can help supplement treatment by remotely monitoring patient symptoms and predicting mental health crises in between clinical visits. However, the feasibility of digital mental health technologies has not yet been sufficiently explored. Rhythms, from the company Health Rhythms, is a smartphone platform that uses passively acquired smartphone data with artificial intelligence and predictive analytics to alert patients and providers to an emerging mental health crisis.

Objective: The objective of this study was to test the feasibility and acceptability of Rhythms among patients attending an academic psychiatric outpatient clinic.

Methods: Our group embedded Rhythms into the electronic health record of a large health system. Patients with a diagnosis of major depressive disorder, bipolar disorder, or other mood disorder were contacted online and enrolled for a 6-week trial of Rhythms. Participants provided data by completing electronic surveys as well as by active and passive use of Rhythms. Emergent and urgent alerts were monitored and managed according to passively collected data and patient self-ratings. A purposively sampled group of participants also participated in qualitative interviews about their experience with Rhythms at the end of the study.

Results: Of the 104 participants, 89 (85.6%) completed 6 weeks of monitoring. The majority of the participants were women (72/104, 69.2%), White (84/104, 80.8%), and non-Hispanic (100/104, 96.2%) and had a diagnosis of major depressive disorder (71/104, 68.3%). Two emergent alerts and 19 urgent alerts were received and managed according to protocol over 16 weeks. More than two-thirds (63/87, 72%) of those participating continued to use Rhythms after study completion. Comments from participants indicated appreciation for greater self-awareness and provider connection, while providers reported that Rhythms provided a more nuanced understanding of patient experience between clinical visits.

Conclusions: Rhythms is a user-friendly, electronic health record–adaptable, smartphone-based tool that provides patients and providers with a greater understanding of patient mental health status. Integration of Rhythms into health systems has the potential to facilitate mental health care and improve the experience of both patients and providers.

(JMIR Form Res 2025;9:e57624) doi:10.2196/57624

KEYWORDS

digital health; mobile intervention; telepsychiatry; artificial intelligence; psychiatry; mental health; depression; mood; bipolar; monitor; diagnostic tool; diagnosis; electronic health record; EHR; alert; notification; prediction; mHealth; mobile health; smartphone; passive; self-reported; patient generated

Introduction

Mental disorders are predicted to be the leading cause of disease burden worldwide by 2030 [1]. However, there is a tremendous mismatch between the number of psychiatrists, psychologists, and other mental health care providers and individuals in need of treatment for mental illness, including substance use disorders. In the United States alone, 160 million people reside where there is a shortage of mental health professionals [2]. Roughly 56% of counties are without a practicing psychiatrist, while 33% of counties are without a licensed psychologist [3]. The workforce is expected to be short 38,821 psychiatrists by 2024 [4]. In the United States, the majority of individuals seeking treatment for depression, anxiety, or any mental illness receive care from their primary care physician (PCPs), and PCPs provide almost a third of the care for serious mental illness [5].

On average, a PCP addresses between 3 to 5 problems, including psychiatric problems, during a typical encounter [6,7]. Furthermore, the majority of PCPs cannot see patients as frequently as may be necessary to adequately monitor a patients' mental illness and response to treatment, nor do they have a mental health specialist to whom they can refer their more concerning patients [8]. In the context of the current mental health landscape, remote monitoring of patient mental well-being between visits is poised to play a major role in the redesigning of mental health care, similar to what is occurring with remote management of myriad other conditions [9,10].

Ideally, remotely acquired data would be integrated into the electronic health record (EHR) and readily available to the patient's provider for interpretation and intervention if needed. With remote patient monitoring, the documentation of history since the previous appointment is quicker, the challenges of patient recall with respect to their signs and symptoms are obviated, and there is a foundation for patients to participate to a greater degree in their own psychiatric care by tracking their symptoms alongside their provider.

Clinical, patient-generated, app-acquired health data that are embedded and presented into the EHR workflow provide the opportunity for interprofessional use. This is particularly important when patients have multiple comorbid medical conditions that impact brain health and vice versa. Furthermore, patient engagement with health apps, including those that are mental health related, declines dramatically over the first few weeks of use, in some cases dropping below 10% within 2 weeks [11,12]. An ideal app may rely less upon active user engagement for data collection and to a greater degree on passively acquired data that reduce patient burden, which then trigger the completion of symptom self-reports using nationally accepted standardized ratings when passively acquired patient-generated health data (PGHD) indicate a relapse is about to occur.

To address these deficits in the digital mental health space, an academic-industry collaborative team was formed to test the feasibility and acceptability of industry start-up Health Rhythms' smartphone solution called Rhythms for use across a large health system. Previous research has demonstrated that Rhythms' proprietary artificial intelligence (AI) algorithm applied to passively collected data indicative of behavioral vital signs (eg, sleep patterns, activity measures, and sociability) can predict a relapse of depression 7 days before the patient's self-report of relapse with 90% accuracy [13-15]. This collaborative team reports the results of the EHR integration, results of the feasibility testing of the urgent and emergent alert algorithms, effectiveness of the 24/7 virtual patient monitoring system, user acceptability in terms of both patient use of and satisfaction with Rhythms, and provider impressions of the impact of Rhythms on patient care.

Methods

Collaborative Team

Health Rhythms is a technology company that was founded by experienced researchers and clinicians (EF and DK) in the assessment, treatment, and research of mental illness, particularly bipolar disorder and depression, in partnership with international leaders (MM) in behavioral sensing technology focused on mental health. The University of Colorado (CU) Anschutz Medical Campus Department of Psychiatry is a large department (500+ faculty) with a strong commitment to the incorporation of digital health and technology in the care of patients with serious mental illness. CU Innovations brings together industry partners, entrepreneurs, and investors to help CU researchers create biomedical technology that improves care delivery, health outcomes, and patient quality of life worldwide. University of Colorado Health (UCHealth) is a not-for-profit health care system across Colorado, southern Wyoming, and western Nebraska with an academic medical center, the University of Colorado Hospital (UCH), that is affiliated with the CU School of Medicine. The CARE Innovation Center at UCHealth partners with CU to create a comprehensive platform of resources both to conduct clinical and biomedical research and to transform real-world clinical practice together with industry partners or faculty entrepreneurs.

App Development

The Rhythms app was developed using a participatory design process [16], an approach that seeks to deeply involve end users in the design process. A weekly meeting between the Health Rhythms team and the UCHealth clinical team was the primary means to engage in this process. It focused on understanding end-user needs, existing clinical workflows, and seeking feedback on early prototypes.

A key decision resulting from this process was to integrate the Health Rhythms software development kit (SDK) into UCHealth's MyChart mobile app, rather than providing a distinct separate app that participants would have to download. Health Rhythms' SDK is a modular piece of software that can be integrated easily into native iOS or Android apps. Although this decision required significant technical work, the team anticipated that it would significantly streamline the patient end-user experience, since they would only have to install and use 1 app, and would help maintain user privacy and security.

Once integrated into the UCHealth MyChart app and end users have provided explicit permissions during onboarding, the SDK collects data continuously in the background from the activity, location, and display smartphone sensors. These data can then be used to automatically assess behavior patterns derived from these data streams including how much time a person is spending at home, levels of physical activity, estimates of sleep timing, and duration and interaction patterns with the smartphone itself.

Safeguarding patient privacy was paramount. Users were uniquely identified using a nonidentifiable identifier that was passed to Health Rhythms. The Rhythms app did not collect any identifiable data that falls under the protections of the Health Insurance Portability and Accountability Act (HIPAA) such as birth dates, home addresses, or social security numbers, and the SDK did not collect any data from texts, calls, photos, internet browsing, or other content. All data, including GPS location, are encrypted using state of the art encryption (ie, a 256-bit key) on the device, in transmission and in the cloud. The sensor data for each patient were computed into daily nonidentifiable behavioral summaries that were then transmitted into the EHR using standard Fast Healthcare Interoperability Resources end points. No unprocessed sensor data, including GPS, were transferred into the EHR.

While there is always a risk of data breach in such systems, the following measures were implemented to mitigate this risk. The team applied the principle of data minimization, reducing the collection, storage, and use of personal data to what is strictly necessary to assess individual behavioral health. Patient end users had the agency to rescind permissions for sensor data collection at any time by a "Manage Settings" screen, and transparent, easy-to-understand messaging during onboarding communicated clearly what data were collected and how they would be used. Finally, Health Rhythms maintained General Data Protection Regulation and California Consumer Privacy Act certification and conducted regular security tests by professional third parties.

AI Methods

The Rhythms system collects passive sensing data, comprising location, pedometer, activity, and device data, from the onboard sensors in smartphones. Each day, these data are processed into 65 behavioral inferences for each person. These inferences encompass measures related to physical and social activity, stability or instability of the person's weekly routine, sleep, and smartphone use. For example, time at home is an inference derived from location data about how much time a person is spending at home.

These behavioral inferences are used in 2 ways. First, heuristics related to these inferences are used to trigger a request for a self-report from participants. A total of 15 static heuristics, empirically derived from previous datasets, dynamically prompted individuals to complete self-reported assessments. For example, if an individual had a diagnosis of major depression and spent a long time at home for over 7 days, then a self-report was triggered. Second, the behavioral inferences were fed into a logistic regression machine learning model that provided a continuous and passive estimate of a person's risk for a relapse in depression, defined as scoring greater than 10 on the Patient Health Questionnaire 9-Item (PHQ-9). This model aims to be predictive of how a given person would answer specific questions on the PHQ-9. Model outputs with a high probability for a high PHQ-9 score were used to flag individuals as a medium alert on the dashboard if there was no corroborating self-report.

Participant Criteria and Recruitment

Potential participants were residents of Colorado; aged between 18 and 89 years; were the primary user of a mobile phone that supported iOS12 or above; enrolled in MyHealthConnection (MHC), which is UCH's EHR patient portal; and were willing to enroll their device into Rhythms.

The UCH's EHR (EPIC) was used to search the caseload of participating clinicians in the CU Department of Psychiatry to identify patients who carry the diagnosis of depressive disorder including major depressive disorder (MDD), dysthymia, depressive disorder due to another medical condition, or bipolar disorder. Providers reviewed the list to confirm that individuals were current patients and would be clinically appropriate for contact about Rhythms.

Recruitment was conducted virtually. Patients were contacted initially through MHC with a general message describing the study. Due to poor initial response to the MHC outreach, a personalized email letter was sent from their provider. Potential participants were contacted up to 3 times by a study coordinator, and screening appointments were scheduled if they expressed interest in the study. After consenting and screening, the study coordinator reviewed how to download the app and interact with the platform and placed an order for Rhythms in EPIC. Participants were considered lost to follow-up if they did not respond to 3 contacts by the study team after enrollment.

Throughout the study, participants provided data by completing electronic surveys sent by the study team (baseline, end of study [6 weeks], and when triggered by algorithm), as well as by active and passive use of the Rhythms platform. A purposively

sampled [17,18] group of participants participated in qualitative interviews with a member of the study team at the end of the study. An overview of the study design and procedures is presented in Figure 1.

Figure 1. Study flow chart: (A) potential participants identified using the electronic health record in collaboration with providers; (B) potential participants contacted up to three times; (C) participants (who met criteria and signed consent) downloaded Rhythms and were instructed in its use; (D) participants completed baseline assessments; (E) participants continued their daily activities, with Rhythms passively collecting data for 6 weeks; (F) participants completed baseline assessment again at end of study; (G) participants completed a survey of likes and dislikes with respect to Rhythms; and (H) selected group underwent qualitative interviews about Rhythms use.

Data Collected by Rhythms

Passively Measured Outcomes

Rhythms acquired proxy information regarding sleep quality, activity level, and sociability. The AI algorithm uses passively collected data to detect deviations from the individual's baseline and triggers a push to complete the self-report ratings described below. During the course of this study, the participant could not view their own passively acquired data.

Self-Report Measures

Per study protocol, self-report measures were administered at baseline and end of study. In addition, self-report measures were pushed by Rhythms to the participant during the study period when the AI algorithm detected a deviation in passively collected data indicative of worsening of symptoms of depression, anxiety, or mania. Participant self-ratings included the following:

PHQ-9 Instrument

The PHQ-9 asks the individual to rate the severity of symptoms over the past 2 weeks on a 4-point scale ranging from "not at all" to "nearly every day." Endorsement of either depressed mood or anhedonia (PHQ-2) triggers further evaluation with the PHQ-9, which assesses for the presence of other criteria for MDD. In multiple studies, PHQ-9 scores >10 had a sensitivity of 88% and a specificity of 88% for MDD [19-21].

Columbia-Suicide Symptom Severity Rating Scale

The Columbia-Suicide Symptom Severity Rating Scale (C-SSRS) is used to assess severity and immediacy of suicide risk using four constructs: (1) severity of suicidal ideation, (2) intensity of ideation, (3) behavior, and (4) lethality. The instrument has been reported to have a 67% sensitivity and 76% specificity for identifying suicidal behaviors [22,23].

Generalized Anxiety Disorder 7-Item

The Generalized Anxiety Disorder 7-Item (GAD-7) is used to measure the severity of generalized anxiety disorder symptoms by asking individuals to rate the severity of symptoms on a 4-point scale with ranging from "not at all" to "nearly every day." Scores of 5, 10, and 15 represent cutoff points for mild, moderate, and severe anxiety, respectively [24].

Altman Self-Rating Mania Scale

The Altman Self-Rating Mania Scale (ASRM) is a 5-item, self-reported scale to assess positive mood, self-confidence, sleep patterns, speech patterns and amount, and motor activity, which are rated on a scale of "0" or "normal" to "4" or "overtly manic." Total scores range from 0 to 20, with scores equal to or higher than 6 indicating a higher likelihood of manic or hypomanic symptoms [25].

Participant-Reported Measures Collected by Study Team

Postintervention Feasibility and Satisfaction Survey

At study end, participants completed a survey about what they liked or disliked most about Rhythms, their interest in continued use, challenges or barriers to use, and any concerns that they may have had regarding alerting and monitoring through the use of Rhythms.

Qualitative Interviews

The qualitative in-depth interviews were conducted by Zoom (Zoom Video Communications) using a semistructured interview guide to ensure key topics were discussed while also allowing for emergence and exploration of topics important to the participants. The main interview questions for patients focused on their perceived benefits from and drawbacks of the Rhythms' system and suggestions for improving Rhythms' features and use.

Providers also completed a feedback survey at study end. Main questions for providers included their experience integrating the Rhythms system into their workflow, how the system was perceived to impact the quality of care, and the efficiency of providing personalized treatment for patients.

Determining Alert Structure

Based upon previous research with Rhythms [14,15,17,26] and consensus of subject matter experts in the treatment of MDD, anxiety, and bipolar disorder, Rhythms was programmed to send urgent or emergent alerts to UCHealth's Virtual Behavioral Health Center (VBHC) or the Department of Psychiatry's care coordinator based upon passively acquired data and participant self-ratings (Textbox 1).

Textbox 1. Urgent and emergent alerts.

Urgent alert

- Physician Health Questionnaire-9 item score is 20-27, when the previous score was lower.
- Columbia-Suicide Severity Rating Scale question 3=yes.
- Altman Self-Rating for Mania Scale score is ≥11.
- Altman Self-Rating for Mania Scale score is <11 and score for question 3=3 or 4.
- If bipolar disorder diagnosis and Rhythms detect a 40% decrease in sleep.
- If bipolar disorder diagnosis and Rhythms monitoring shows no sleep in 2 days.
- Generalized Anxiety Disorder-7 item score ranges from 15-21, when the previous score was not in this range.

Emergent alert

- Columbia-Suicide Severity Rating Scale questions 4, 5, or 6=yes.
- Altman Self-Rating for Mania Scale total score >11 and answer to question 3=4.

The following specific self-rating questions pertained to the alert structure: C-SSRS question 3: Active suicidal ideation with any methods (not plan) without intent to act; C-SSRS question 4: Active suicidal ideation with some intent to act, without specific plan; C-SSRS question 5: Active suicidal ideation with specific plan and intent; C-SSRS question 6: Have you ever done anything, started to do anything, or prepared to do anything to end your life?; and ASRM question 3: Regarding Sleep Patterns: score of 3="I frequently need less sleep than usual." and score of 4="I can go all day and night without any sleep and not feel tired."

Per protocol, a care coordinator from the VBHC or the Department of Psychiatry called the participant up to 3 times within 24 hours in response to an *urgent* alert. The VBHC monitor covered alerts over weekends, holidays, and after hours, while the care coordinator supporting the participant's primary mental health care provider covered during regular work hours. When the participant was reached by the VBHC, a brief safety assessment was conducted; participants were offered the chance to speak to an on-call clinician for the VBHC and referred to their current provider if appropriate. In the case of clinic care coordinators, the participant was assessed with respect to safety, possible medication changes, and need for earlier appointment with their provider.

In response to *emergent* alerts, Health Rhythms sent a page to the VBHC monitor. The pager beeped every 5 minutes until the alert was acknowledged by the monitor in the EHR. All emergent alerts were managed by the VBHC monitor with the expectation that a call to the patient would be initiated within 15 minutes of the alert. Upon reaching the patient, a safety assessment was conducted, and an appropriate disposition was managed. If after 3 attempts there was no contact with the patient, the VBHC monitor called emergency services to conduct a welfare check at the address listed for the participant.

Initially, participants received a message each time they completed self-report measures that they could receive a call from the VBHC based upon how they answered the questionnaires. This caused confusion for participants and the process was changed such that only participants who had triggered an emergent or urgent alert would get the message that they would receive a call.

Data Analysis

Demographics and baseline behavioral rating scores as well as number and nature of alerts were analyzed descriptively with means (SDs) and frequencies (percentages). Postintervention survey data were summarized descriptively for fixed-choice items. Open-ended survey items and interview transcripts were analyzed qualitatively. Human-adjudicated transcription of interview recordings was performed with Otter.ai, a proprietary AI-powered note taking engine, to generate initial transcripts, after which a member of the study team reviewed the transcripts while simultaneously listening to the recordings to ensure there were no errors in transcription.

A rapid qualitative analysis approach was used for content analysis of open-ended survey and interview transcript data. We first deductively explored and identified common themes that appeared in participants' responses. Two members of the research team separately reviewed the data and identified keywords and topics that are relevant to our research questions (eg, most liked and disliked features of the Rhythms' system, facilitators and barriers to using the Rhythms system, and suggestions to improve the system). During this process, each reviewer independently summarized how frequently common items were reported by study participants. The reviewers compared their results, discussed and resolved any discrepancies, reached consensus on the latent themes that had emerged in analysis, and generated a summary.

Ethical Considerations

This project was approved by the University of Colorado Multiple Institution Review Board (COMIRB; 21-4903), and participants gave electronic informed consent. Participants received US $25 in compensation for completing the screening, enrollment, and baseline surveys, and an additional US $50 for completing the postintervention surveys as described below. All data were deidentified for analysis and presented anonymously.

Results

Participants

Of 394 patients who were informed about the study through MHC or email, 178 (45.2%) did not respond, 26 (6.7%) were contacted but lost to follow-up before enrollment, 86 (21.8%) were screened but did not enroll, and 104 (26.4%) were enrolled within 8 weeks of study initiation. The most common reason (60/86, 69.8%) for not meeting study criteria was having a phone using the Android operating system.

The study sample comprised 104 individuals; the majority identified as women (n=72 69.2%), non-Hispanic (n=100 96.2%), and White (n=84 80.8%). The average age was 42.1 (SD 15.9) years. The majority of participants had MDD (n=71 68.3%), while 23 (22.2%) individuals had either bipolar I or II disorder and 10 individuals (9.6%) had “other mood disorder” (Table 1).

Between enrollment and study completion, 17 (16.3%) of the 104 enrolled participants were lost to follow-up, withdrew from the study, or did not complete the final survey. Of note, among the 87 (83.7%) completers out of 104, a total of 72% (63/87) continued using Rhythms despite no longer being compensated for doing so.

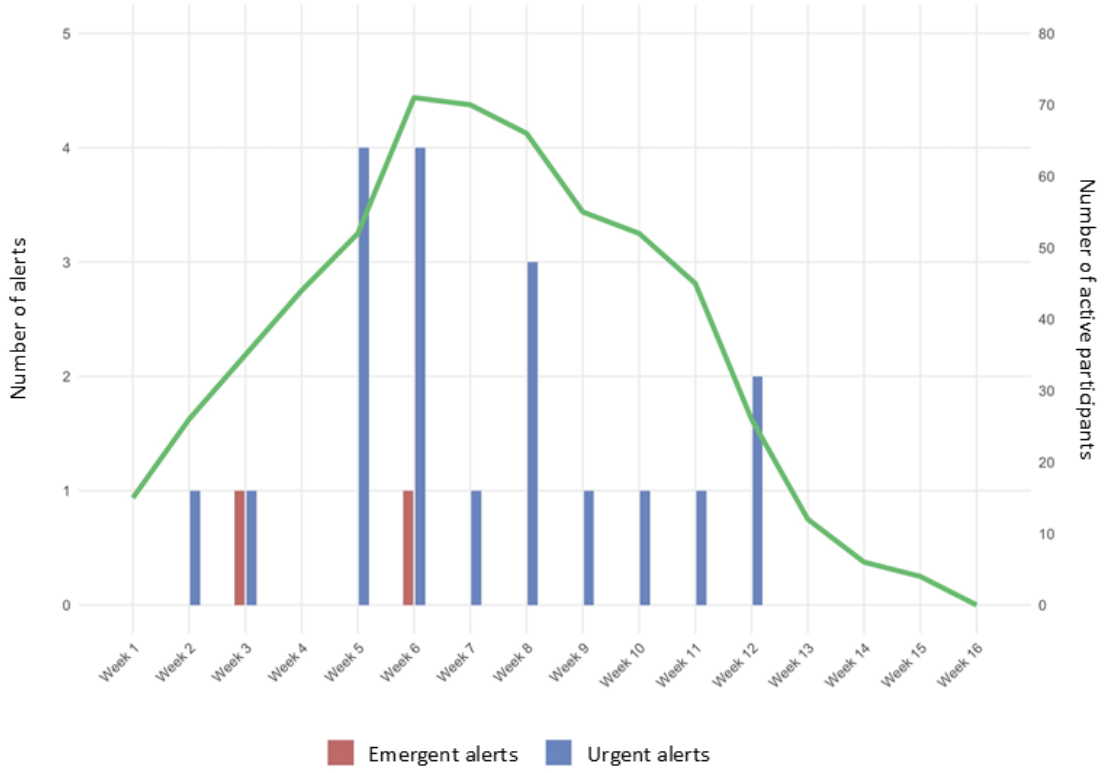
Table 1. Participant characteristics and baseline behavioral ratings.

Variable	Value (n=104)
Age (years), mean (SD)	42.1 (15.9)
Gender, n (%)	
Women	72 (69.2)
Men	29 (27.9)
Transgender or nonbinary	3 (2.9)
Ethnicity, n (%)	
Hispanic	3 (2.9)
Non-Hispanic	100 (96.2)
Not answered	1 (1)
Race, n (%)	
Asian	9 (8.6)
Black or African American	1 (1)
American Indian or Alaska Native	1 (1)
White	84 (80.8)
More than one race	4 (3.8)
Other or prefer not to answer	5 (4.8)
Diagnosis, n (%)	
Major depressive disorder	71 (68.3)
Bipolar 1 disorder	12 (11.5)
Bipolar II disorder	11 (10.6)
Other mood disorder	10 (9.6)
Ratings, mean (SD)	
Physician Health Questionnaire 9-Item	6.5 (5.3)
Generalized Anxiety Disorder 7-Item	5.9 (4.8)
Altman Self-Rating Mania Scale	2.2 (2.7)

Urgent and Emergent Alerts

Over the course of the entire 16-week study period, the monitoring of 104 individuals with Rhythms for 6 weeks each resulted in 2 emergent alerts and 19 urgent alerts. The largest number of participants being monitored at 1 time was 92 and the average was 64. The greatest number of alerts on any given day was 3 (all urgent, 1 day only) when there were 58 participants being monitored (Figure 2).

Figure 2. Urgent and emergent alerts across time and by number of active participants.



The first emergent alert was triggered by a participant answering ASRM question 3 about sleep with a "4," indicating "I can go all day and night without any sleep and not feel tired." The second emergent alert was triggered by the participant answering yes to C-SSRS question 6, indicating that they "had ever done anything, started to do anything or prepared to do anything to end their life." The primary (n=15) trigger for urgent alerts was answering "yes" to question 3 on the C-SSRS, indicating that the participant had active suicidal ideation without method or intent to act. The second most common (n=12) reason for urgent alerts was a GAD-7 score indicating severe anxiety. Three urgent alerts occurred because of the ASRM indicating mania or hypomania based upon total ASRM or response to question 3 regarding need for sleep.

All calls to participants were managed within the specified alert time. Eleven (58%) of the 19 urgent alerts came through during work hours and were managed by the Department of Psychiatry's care coordinator. The 2 emergent alerts were managed by the VBHC, neither of which required an emergency room referral or in-person well-being check.

Postintervention Survey: Patients and Providers

Of the 87 participants who completed the postintervention feasibility and satisfaction survey, the median score for ease of starting the system, ease of use, and surveys being delivered at convenient times were 9, 10, and 9, respectively, on a scale of 0 (strongly disagree) to 10 (strongly agree). Median score for participant satisfaction with the Rhythms system within their MHC app was 8.

Themes from the free text indicated that respondents most liked that Rhythms gave them a greater awareness of their own mental health status, the feeling of being taken care of by their provider, and its ease of use. The most common complaint about Rhythms (mentioned by n=18, 21%) was having to keep the app open. Only 3 (3%) reported concerns about being "tracked."

When providers were surveyed about their experience with Rhythms, they noted that the platform provided them with information they would not otherwise have had and information that was directly actionable. For example, 1 provider commented that "...the sleep section was the most useful for me to review. I am very surprised that one of my patients is only averaging about 5 hours of sleep per night, so that is definitely something we will talk about." Another commented that "The more frequent assessment was great. I don't think patients remember their condition for more than a few days. I was able to reach out to a couple of patients based on worsening in [their] data."

Qualitative Interviews: Patients

Qualitative interviews with 19 (22%) of the 87 completers revealed that all but 2 were well established in the CU and UCH system for their mental health care, having received care there for at least 3 years. Most agreed to participate in the study because they felt positively about research (n=10, 11%), thought it would be easy (n=4, 5%), or thought it would be helpful to either themselves (n=3, 3%) or others (n=3, 3%). Setup and initial use of Rhythms was reported by 13 (68%) of those interviewed as being "simple, intuitive, and unobtrusive," although 5 (26%) reported some technical challenges, including difficulty with changing settings, granting permissions, or finding specific features. With respect to completing the self-report measures and their frequency, 63% (n=12) of respondents reported completing them at the time of the prompt. In general, respondents were okay with the frequency of measures (n=3, 3%) or agreed that weekly (n=3, 3%) or every 2 weeks (n=4, 5%) would be acceptable, although a few (n=2,

2%) preferred monthly or felt surveys were sent too often (n=1, 1%).

Overall, interview participants reported positive perceptions of Rhythms, finding it easy to use (n=8, 9%) and enjoying its use (n=5, 6%). With respect to what was liked most about Rhythms, there was consensus that the platform promoted greater provider connection (n=4, 5%) and provider awareness of patients' health status (n=6, 7%), contributing to a feeling of being "looked out for" (n=4, 5%). Participants also reported improved self-awareness of one's own mental health status (n=5, 6%). What was liked least about Rhythms was having to keep the app open (n=7, 8%), with related concern for battery life (n=4, 5%); the wording of a response message to completed surveys notifying patients that they might be contacted based on their responses (n=6, 7%); concerns about what would happen if answers to questions were negative (n=3, 3%); not receiving immediate feedback from Rhythms; and accuracy of the data if the person did not have their phone with them or were traveling outside of Colorado.

Roughly half of those interviewed endorsed having had a helpful discussion about their Rhythms data with their provider, while others indicated that their provider did not mention their data during their appointment, or they did not have an appointment during the 6 weeks of tracking. The most common recommendation for system improvement was the ability to access and track their own data in a visual format, an improvement that is now available to users. Illustrative quotations about positive aspects of Rhythms were as follows:

> *…I had one two-week period where it was like, everything's catastrophic…which solicited the call from [my provider], through the patient portal. "So I saw your responses. What's, the scoop?"…I liked the fact that somebody was looking out for me. It just—I'm, I'm an introvert. I don't want to bother people. You know, am I really so bad that I need to interrupt somebody today? I don't want to be a bother. But I was able to experience the effectiveness of the system, because [my provider] reached out to me. So, from a patient point of view, I can tell you, that particular item worked.*

> *And then there was one survey when I was starting to feel a little more anxious than I normally do. I got asked to complete a survey. And I did. And I know when I completed that survey, [my provider] reached out to me and said, "Hey, I noticed that you're feeling a little more anxious than you seem to be normally." So, I was. I'm not sure what prompted that. But I thought it was interesting that at a time I kind of had a need to communicate with [my provider], the system seemed to have recognized that.*

> *Overall, I think that it is a wonderful tool to have in one's toolbox to help support mental wellness.*

Illustrative quotations regarding concern about Rhythms were as follows:

> *Also, I remember first time and a few times after that, too, I got - because I know sometimes when you answer things in a way that is concerning, they say someone will call after every survey, I got that message. Sometimes it was someone may call and sometimes someone will call…I remember because one time I left my phone at home and I went out and I was like, oh my god, someone's gonna call and they're gonna message. Yeah, that message was a little scary sometimes…*

> *I didn't like hearing that I may or may not receive a call from a provider, I would have liked it if it was more definitive.*

> *Like, I don't sleep with my mobile device…and I don't always exercise with my phone.*

Reliability

The Rhythms system performed well during the pilot study with no outages and no major bugs detected.

Discussion

Principal Findings

Here, we describe the use, including acceptability, of the smartphone platform Rhythms by patients and providers in an academic department of psychiatry. Overall, the system integrated well with the health system's EHR and clinical workflow and was able to be deployed with very few technical issues to a group of 104 participants. The algorithm for urgent and emergent alerts allowed the team to identify concerning changes in patient mental health status and may be considered by other clinical groups wishing to use Rhythms. The lack of adverse outcomes during this period suggests that the algorithm was set conservatively enough to capture those who are in need of outreach, but not so sensitive as to lead to inappropriate alerts that could unnecessarily burden the patient or overwhelm the monitoring team.

An impressive percentage (63/87, 72%) of study participants remained on the platform and continued to use it without any compensation beyond the end of the 6-week study, suggesting they found the app helpful. This is in distinction to the findings from Kopka and colleagues [11] showing that use of 4 out of 10 popular mental health apps dropped below 20% within the 2 weeks of initiation. None of the 10 apps chosen by participants in Kopka and colleagues [11] collected passively acquired data or were integrated into the individual's health record. In our study, poststudy interviews revealed that passively collected data lead to a greater connection with one's provider and was a major strength of Rhythms. Participants also provided important feedback regarding the functionality (ie, ability to track ratings visually over time and receiving immediate feedback) of Rhythms that has now been incorporated into the platform. Importantly, privacy was not a major concern expressed by participants in this study, unlike those in some [26-29] but not all previous studies [12].

The concern about having to leave the app open is real, but this did not deter the majority of individuals from remaining on the app. While providers reported that Rhythms is a useful tool in their clinical practice, some participants reported that their provider did not discuss their Rhythms data during their

appointment. This may reflect the timing of Rhythms use and the availability of collected data in relation to participants' appointments, or it may be indicative of differing levels of provider adoption of the technology and how they incorporated it into their own practices. The use of objective measurements for patient care, which is quite common in most fields of medicine, is only recently making its way into the practice of psychiatry. We are encouraged that when the Rhythms data were brought into the session, the patients reported a benefit. While scalable technology holds great promise for improving psychiatric treatment, the acceptance and use of technology in psychiatry would require a culture change if measurement assisted care is to become a reality. Since the COVID-19 pandemic, many psychiatrists and other mental health specialists are now comfortable using videoconferencing for patient visits. However, incorporating data collected from apps and wearables is an unfamiliar step in the therapeutic interaction.

Likewise, many health systems have not yet embraced linkage of apps used to collect PGHD for mental health care purposes with their EHR. When there is collection of these data through medical records standards initiative such as Fast Healthcare Interoperability Resources and other platforms [30], information is not easily collated across multiple patient-derived sources. Recent advances have been made in this area through platforms, such as mindLAMP from the Digital Psychiatry Program at Beth Israel Deaconess Medical Center of Harvard Medical School, which has the power to incorporate PGHD collected by multiple apps and wearables and offers tools to help configure data such that they are more readily usable for research or clinical care [31]. Within this context, the benefit of Rhythms is that it can be seamlessly integrated into the patient's EHR such that the provider can view passively acquired data as well as ratings with 1 click in the patient's chart. The alert algorithm is strong and can help guide others in their use of Rhythms to detect emerging patient crises.

Those psychiatrists who incorporated Rhythms data into their clinical care reported that it improved their understanding of their patients' clinical status and that they used these data in their treatment planning. Likewise, patients appreciated the added connection to their provider, and there were no reports of increased provider burden across the study.

Several limitations should be considered. In total, 45.2% (178/394) of potential participants did not respond to either a general inquiry sent through MHC or an email letter sent by their provider. It is unclear why they did not respond to messages through the EHR, although previous research suggests that patient preferences for communication method are influenced by the type of information to be received [32]. Willingness to participate in a study that requires downloading an app onto one's personal phone may require direct interaction with one's provider and a clear discussion of the potential benefits for their treatment. The sample predominantly identified as women and White, limiting generalizability of these findings to other populations. The relatively short duration of this initial trial (6 weeks for each participant) and the relatively small number of patient and provider participants was also a limitation. Though the clinical value of Rhythms was quickly apparent, a longer period of monitoring is needed to determine the economic impact of remote patient monitoring of mental health status on health care use and costs. All study participants were currently in active mental health treatment and, on the whole, relatively stable. How Rhythms would perform in a more acutely ill population or those with psychosis within the greater health system is unknown but should be considered in future studies, as psychiatric disorders can negatively impact one's motivation and ability to engage with an app and their overall health care [33,34].

In summary, Rhythms is a user-friendly, digital platform that can be embedded into a health system's EHR and used to identify patients who are experiencing a worsening of their mental health between visits with their mental health care provider. The AI algorithms are based upon passively acquired data, limiting the need for patient surveys except when deviations in data suggest a deterioration in the patient's clinical status. The culture surrounding the use of technology as part of measurement assisted mental health care will become more welcoming as studies such as ours show the ease and clinical benefits of using remote patient monitoring tools.

Acknowledgments

This work was funded in part by a research grant from Health Rhythms. The authors thank Dr. Chloe E Page for assistance with submission and copyediting.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

CNE contributed to the study conception and design, data collection, analysis and interpretation of results, and manuscript preparation. RD contributed to study design, data collection, and manuscript preparation. AD contributed to data collection, interpretation of results, and manuscript preparation. HCH contributed to data collection, data management, and manuscript preparation. DJK participated in study conception and manuscript preparation. TL contributed to study design, data collection, and manuscript preparation. PMV participated in study design, data collection, and manuscript preparation. MM participated in data collection and manuscript preparation. SLM participated in study conception and design, data collection, analysis and interpretation of results, and manuscript preparation. KM participated in study conception and preparation. CDS participated in

study design, data collection, and manuscript preparation. JLS participated in study design, data collection, analysis and interpretation of results, and manuscript preparation. RDZ contributed to manuscript preparation. EF contributed to study conception and design and manuscript preparation.

Conflicts of Interest

CNE is a consultant to Skyland Trail, BabyScripts, and EmbarkNeuro. CNE was the principal investigator on this study, which was supported in part by a grant from Health Rhythms. EF is an employee of and holds equity in Health Rhythms. TL, MM, CNE, and DJK hold equity in Health Rhythms. AD received research grant support from Health Rhythms. KM is an investor in Health Rhythms.

References

1. Arias D, Saxena S, Verguet S. Quantifying the global burden of mental disorders and their economic value. EClinicalMedicine 2022;54:101675 [FREE Full text] [doi: 10.1016/j.eclinm.2022.101675] [Medline: 36193171]
2. Counts N. Understanding the US behavioral health workforce shortage. The Commonwealth Fund. 2023. URL: https://www.commonwealthfund.org/publications/explainer/2023/may/understanding-us-behavioral-health-workforce-shortage [accessed 2024-11-02]
3. Beck AJ, Page C, Buche J, Rittman D, Gaiser M. Estimating the distribution of the U.S. psychiatric subspecialist workforce. Health Workforce Technical Assistance Center. 2018. URL: https://www.healthworkforceta.org/research-alerts/estimating-the-distribution-of-the-us-psychiatrist-subspecialist-workforce/ [accessed 2024-11-08]
4. Satiani A, Niedermier J, Satiani B, Svendsen DP. Projected workforce of psychiatrists in the United States: a population analysis. Psychiatr Serv 2018;69(6):710-713. [doi: 10.1176/appi.ps.201700344] [Medline: 29540118]
5. Jetty A, Petterson S, Westfall JM, Jabbarpour Y. Assessing primary care contributions to behavioral health: a cross-sectional study using medical expenditure panel survey. J Prim Care Community Health 2021;12:21501327211023871 [FREE Full text] [doi: 10.1177/21501327211023871] [Medline: 34109860]
6. Kyanko KA, A Curry L, E Keene D, Sutherland R, Naik K, Busch SH. Does primary care fill the gap in access to specialty mental health care? a mixed methods study. J Gen Intern Med 2022;37(7):1641-1647 [FREE Full text] [doi: 10.1007/s11606-021-07260-z] [Medline: 34993864]
7. Beasley JW, Hankey TH, Erickson R, Stange KC, Mundt M, Elliott M, et al. How many problems do family physicians manage at each encounter? a WReN study. Ann Fam Med 2004;2(5):405-410 [FREE Full text] [doi: 10.1370/afm.94] [Medline: 15506571]
8. Richman EL, Lombardi BM, Zerden LD. Mapping colocation: using national provider identified data to assess primary care and behavioral health colocation. Fam Syst Health 2020;38(1):16-23. [doi: 10.1037/fsh0000465] [Medline: 32202831]
9. Drincic A, Prahalad P, Greenwood D, Klonoff DC. Evidence-based mobile medical applications in diabetes. Endocrinol Metab Clin North Am 2016;45(4):943-965 [FREE Full text] [doi: 10.1016/j.ecl.2016.06.001] [Medline: 27823614]
10. Su D, Michaud TL, Estabrooks P, Schwab RJ, Eiland LA, Hansen G, et al. Diabetes management through remote patient monitoring: the importance of patient activation and engagement with the technology. Telemed J E Health 2019;25(10):952-959. [doi: 10.1089/tmj.2018.0205] [Medline: 30372366]
11. Kopka M, Camacho E, Kwon S, Torous J. Exploring how informed mental health app selection may impact user engagement and satisfaction. PLOS Digit Health 2023;2(3):e0000219 [FREE Full text] [doi: 10.1371/journal.pdig.0000219] [Medline: 36989237]
12. 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;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
13. Frank E, Wallace ML, Matthews MJ, Kendrick J, Leach J, Moore T, et al. Personalized digital intervention for depression based on social rhythm principles adds significantly to outpatient treatment. Front Digit Health 2022;4:870522 [FREE Full text] [doi: 10.3389/fdgth.2022.870522] [Medline: 36120713]
14. Malkoff-Schwartz S, Frank E, Anderson BP, Hlastala SA, Luther JF, Sherrill JT, et al. Social rhythm disruption and stressful life events in the onset of bipolar and unipolar episodes. Psychol Med 2000;30(5):1005-1016. [doi: 10.1017/s0033291799002706] [Medline: 12027038]
15. Abdullah S, Murnane ELM, Matthews M, Choudhury T. Circadian computing: sensing, modeling, and maintaining biological rhythms. In: Rehg J, Murphy S, Kumar S, editors. Mobile Health: Sensors, Analytic Methods, and Applications. Cham, Switzerland: Springer; 2017:35-58.
16. Schuler D, Namioka A. Participatory Design: Principles and Practices. Boca Raton, FL: CRC press; 1993.
17. Patton MQ. Qualitative Research and Evaluation Methods. 3rd ed. Thousand Oaks, CA: Sage Publications; 2002.
18. Creswell JW, Clark VLP. Designing and Conducting Mixed Method Research. 2nd ed. Thousand Oaks, CA: Sage Publications; 2011.
19. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]

20. Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. Psychiatric Annals 2002;32(9):509-515. [doi: 10.3928/0048-5713-20020901-06]
21. Gilbody S, Richards D, Brealey S, Hewitt C. Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): a diagnostic meta-analysis. J Gen Intern Med 2007;22(11):1596-1602 [FREE Full text] [doi: 10.1007/s11606-007-0333-y] [Medline: 17874169]
22. 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;168(12):1266-1277 [FREE Full text] [doi: 10.1176/appi.ajp.2011.10111704] [Medline: 22193671]
23. Mundt JC, Greist JH, Jefferson JW, Federico M, Mann JJ, Posner K. Prediction of suicidal behavior in clinical research by lifetime suicidal ideation and behavior ascertained by the electronic Columbia-Suicide Severity Rating Scale. J Clin Psychiatry 2013;74(9):887-893. [doi: 10.4088/JCP.13m08398] [Medline: 24107762]
24. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
25. Altman EG, Hedeker D, Peterson JL, Davis JM. The Altman Self-Rating Mania Scale. Biol Psychiatry 1997;42(10):948-955. [doi: 10.1016/S0006-3223(96)00548-3] [Medline: 9359982]
26. Matthews M, Abdullah S, Gay G, Choudhury T. Tracking mental well-being: balancing rich sensing and patient needs. Computer 2014;47(4):36-43. [doi: 10.1109/mc.2014.107]
27. Melcher J, Camacho E, Lagan S, Torous J. College student engagement with mental health apps: analysis of barriers to sustained use. J Am Coll Health 2022;70(6):1819-1825. [doi: 10.1080/07448481.2020.1825225] [Medline: 33048626]
28. 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;5(3):e56 [FREE Full text] [doi: 10.2196/mental.9539] [Medline: 30158102]
29. 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;7(4):e11223 [FREE Full text] [doi: 10.2196/11223] [Medline: 30990458]
30. Hassol A, Walker JM, Kidder D, Rokita K, Young D, Pierdon S, et al. Patient experiences and attitudes about access to a patient electronic health care record and linked web messaging. J Am Med Inform Assoc 2004;11(6):505-513 [FREE Full text] [doi: 10.1197/jamia.M1593] [Medline: 15299001]
31. Mandel JC, Kreda DA, Mandl KD, Kohane IS, Ramoni RB. SMART on FHIR: a standards-based, interoperable apps platform for electronic health records. J Am Med Inform Assoc 2016;23(5):899-908 [FREE Full text] [doi: 10.1093/jamia/ocv189] [Medline: 26911829]
32. Vaidyam A, Halamka J, Torous J. Enabling research and clinical use of patient-generated health data (the mindLAMP Platform): digital phenotyping study. JMIR Mhealth Uhealth 2022;10(1):e30557 [FREE Full text] [doi: 10.2196/30557] [Medline: 34994710]
33. 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;21(3):116-119 [FREE Full text] [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
34. Stiles-Shields C, Montague E, Lattie EG, Kwasny MJ, Mohr DC. What might get in the way: barriers to the use of apps for depression. Digit Health 2017;3:2055207617713827 [FREE Full text] [doi: 10.1177/2055207617713827] [Medline: 29942605]

Abbreviations

AI: artificial intelligence
ASRM: Altman Self-Rating Mania Scale
C-SSRS: Columbia-Suicide Symptom Severity Rating Scale
COMIRB: University of Colorado Multiple Institution Review Board
CU: University of Colorado
EHR: electronic health record
GAD-7: Generalized Anxiety Disorder 7-Item
HIPAA: Health Insurance Portability and Accountability Act
MDD: major depressive disorder
MHC: MyHealthConnection
PCP: primary care physician
PGDH: patient-generated health data
PHQ-9: Patient Health Questionnaire 9-Item
SDK: software development kit
UCH: University of Colorado Hospital
UCHealth: University of Colorado Health
VBHC: Virtual Behavioral Health Center

Edited by A Mavragani; submitted 21.02.24; peer-reviewed by D King, PH Liao; comments to author 22.06.24; revised version received 23.07.24; accepted 06.09.24; published 07.01.25.

Please cite as:
Epperson CN, Davis R, Dempsey A, Haller HC, Kupfer DJ, Love T, Villarreal PM, Matthews M, Moore SL, Muller K, Schneck CD, Scott JL, Zane RD, Frank E
The Trifecta of Industry, Academic, and Health System Partnership to Improve Mental Health Care Through Smartphone-Based Remote Patient Monitoring: Development and Usability Study
JMIR Form Res 2025;9:e57624
URL: https://formative.jmir.org/2025/1/e57624
doi:10.2196/57624
PMID:

Original Paper

Co-Design of an Escape Room for e-Mental Health Training of Mental Health Care Professionals: Research Through Design Study

Joyce J P A Bierbooms[1], PhD; Wouter R J W Sluis-Thiescheffer[2], PhD; Milou Anne Feijt[3], MSc; Inge M B Bongers[1], PhD

[1]Tilburg University, Tilburg, Netherlands

[2]HAN University of Applied Sciences, Arnhem and Nijmegen, Netherlands

[3]Eindhoven University of Technology, Eindhoven, Netherlands

Corresponding Author:
Joyce J P A Bierbooms, PhD
Tilburg University
PO box 90153
Tilburg, 5000 LE
Netherlands
Phone: 31 13 466 31 39
Email: j.j.p.a.bierbooms@tilburguniversity.edu

Abstract

Background: Many efforts to increase the uptake of e-mental health (eMH) have failed due to a lack of knowledge and skills, particularly among professionals. To train health care professionals in technology, serious gaming concepts such as educational escape rooms are increasingly used, which could also possibly be used in mental health care. However, such serious-game concepts are scarcely available for eMH training for mental health care professionals.

Objective: This study aims to co-design an escape room for training mental health care professionals' eMH skills and test the escape room's usability by exploring their experiences with this concept as a training method.

Methods: This project used a research through design approach with 3 design stages. In the first stage, the purpose, expectations, and storylines for the escape room were formulated in 2 co-design sessions with mental health care professionals, game designers, innovation staff, and researchers. In the second stage, the results were translated into the first escape room, which was tested in 3 sessions, including one web version of the escape room. In the third stage, the escape room was tested with mental health care professionals outside the co-design team. First, 2 test sessions took place, followed by 3 field study sessions. In the field study sessions, a questionnaire was used in combination with focus groups to assess the usability of the escape room for eMH training in practice.

Results: An escape room prototype was iteratively developed and tested by the co-design team, which delivered multiple suggestions for adaptations that were assimilated in each next version of the prototype. The field study showed that the escape room creates a positive mindset toward eMH. The suitability of the escape room to explore the possibilities of eMH was rated 4.7 out of 5 by the professionals who participated in the field study. In addition, it was found to be fun and educational at the same time, scoring 4.7 (SD 0.68) on a 5-point scale. Attention should be paid to the game's complexity, credibility, and flexibility. This is important for the usefulness of the escape room in clinical practice, which was rated an average of 3.8 (SD 0.77) on a 5-point scale. Finally, implementation challenges should be addressed, including organizational policy and stimulation of eMH training.

Conclusions: We can conclude that the perceived usability of an escape room for training mental health care professionals in eMH skills is promising. However, it requires additional effort to transfer the learnings into mental health care professionals' clinical practice. A straightforward implementation plan and testing the effectiveness of an escape room on skill enhancement in mental health care professionals are essential next steps to reach sustainable goals.

(JMIR Form Res 2025;9:e58650) doi:10.2196/58650

KEYWORDS

serious gaming; mental health care professionals; e-mental health; skill enhancement; training

Introduction

Background

Despite a growing body of evidence for the effectiveness and benefits of e-mental health (eMH), many studies report a relatively low use of eMH in practice [1-3]. eMH means mental health services delivered or supported by digital technologies, such as video calling, self-management and self-tracking apps, web-based treatment modules, and virtual reality (VR) [4,5]. An important factor hampering the use of eMH is its relatively slow adoption by mental health care professionals [5-7]. To a large extent, this slow adoption results from a lack of knowledge and skills regarding eMH among mental health care professionals [7-9].

This lack of skills entails general digital skills but, more importantly, the ability to use an adequate communication approach compensating for a potential lack of nonverbal cues and contextual information. In addition, skills are needed to guide a professional in handling boundaries (eg, expectations regarding reaction time and attainability), choosing appropriate communication channels for each situation, and up-to-date knowledge of the availability of technological possibilities to enrich treatment in mental health care [10-12]. The COVID-19 pandemic caused a significant increase in video calling and, subsequently, the skills to apply such technologies for therapeutic purposes. However, mental health care professionals are still hesitant to pursue broader incorporation of eMH (ie, using different eMH tools in a context where remote treatment is a choice rather than an imposition) into their therapeutic practices because of the skills required [13]. Thus, to increase the adoption of a broad range of eMH tools, there is a persistent need to enhance mental health care professionals' skills in using eMH.

Many training options have recently been developed to capacitate mental health care professionals in using eMH [14,15]. The introduction and development of web-based treatment platforms have brought different programs to acquire the technical skills to use such a platform. Examples are applying web modules, written client interactions, and video sessions instead of real-life therapeutic interactions. Besides the classical training approach, there are several eLearning options to support mental health care professionals in digital skill development [16]. However, research on the effectiveness of different learning and training strategies in health care has shown that traditional forms, such as classical instructions and eLearning, generate insufficient learning outcomes [17,18]. This can be explained by a lack of experiential learning in conventional learning methods; there is a need for a "learning by doing" approach. The learning by doing approach is more engaging, more realistic, and contributes to a larger self-efficacy [17,18]. Thus, there is a need to pursue new, more innovative, and experiential solutions to train mental health care professionals in using eMH.

To gain a feeling of self-efficacy regarding the application of eMH, a potential strategy is to offer mental health care professionals training possibilities based on the concept of serious gaming [19]. Serious games are physical or digital games applied to learn or acquire skills and usually offer a combination of educational content and engagement [19,20]. Escape rooms have emerged as a popular instructional approach within serious gaming, attracting growing interest from researchers and educators [21]. An escape room is a game in which participants are confronted with a storyline, usually locked into a room, and need to solve puzzles, search for information, and follow leads to free themselves from the room within a limited amount of time. Escape room games are increasingly used for learning [22-24].

The benefit of an escape room may be that it offers the option to train skills by gaining hands-on experiences (experiential learning). Escape rooms create realistic learning environments that foster meaningful participation and provide a supportive atmosphere for experimentation [25]. In contrast, other methods (eg, classical instructions or eLearning) mainly focus on providing instructions on using an artifact or service [18,26]. In addition, an escape room is focused on the learning process while simultaneously being educational and fun, which makes it much more engaging to participate in training [27,28].

The underlying learning mechanisms of an educational escape room (EER) are often described in terms of the self-determination theory (SDT) [29,30] and the flow theory [30,31]. The SDT refers to fundamental psychological needs such as autonomy, competence, and relatedness to others [29,30]. EERs provide a safe environment that can be tuned to the player's competencies, and an appropriate storyline helps players relate the competencies demanded in the EER puzzles to competencies required in real-world situations. Together, they increase a player's sense of autonomy, and they can create a positive experience in self-determination. The flow theory explains the occurrence of someone's complete immersion into a specific task based on positive experiences [30,31]. This "flow" can be achieved in serious gaming if the challenges meet the player's knowledge and abilities. By attempting to address these fundamental needs in a playful environment, users can have a positive learning experience [32].

Theoretically, these mechanisms provide a solid argument for the use of EERs to train mental health care professionals' skills in successfully integrating eMH into their therapeutic practice [22,23]. However, as Vorderobermeier et al [30] indicate, the success of EERs is hardly ever tracked back to these mechanisms. To contribute to the emerging field of research into EERs, we include an evaluation of the participants' learning experience in relation to these learning mechanisms.

EERs in eMH

In eMH, the concept of EERs is recognized as a safe and interactive environment to explore eMH in different possible therapeutic situations [30,32] while providing a safe and social environment for training in mental health care. EERs provide

an opportunity to experiment with otherwise complex or expensive therapeutic situations [33,34]. As EERs are an emerging phenomenon, there are also some recent examples of research into EERs in the context of eMH.

In the context of raising awareness of severe mental illnesses (SMIs), Rodriguez-Ferrer et al [35] studied the effectiveness of an EER in educating first-year nursing students (a nonspecialized audience) in experiencing the stigmas associated with SMI. The EER was developed with professionals in SMI and the effect of raising awareness was tested using a randomized controlled trial experiment. Apart from the primary objective (raising awareness of SMI), the EER objectives were to determine whether an entirely web-based version was effective and whether participants were sufficiently immersed in the experience. The EER in this study is different in the sense that it is aimed at experiential learning, involving both digital and physical interactions. Moreover, the EER developed in this study is aimed at a professional audience, which means that the challenges are tuned toward expert knowledge and an in-depth skillset rooted in ample experience in treating mental illnesses [35].

Aragon [36] designed and studied the use of an EER for mental health nursing students attending an accelerated associate degree program. The topic of learning was communication and collaboration strategies in mental health. The design of the EER took a pragmatic, digital form to simplify the strategy for the development, distribution, and deployment of the EER in the educational program. The authors studied the potential learning effect of the digital EER but found no significant difference compared to students who were not exposed to the EER. The EER in this paper is different, similar to the EER studied by Rodriguez-Ferrer et al [35] mentioned earlier, as it targets a professional audience and aims to provide a more physical, collaborative learning experience. Moreover, the EER by Aragon [36] was solely designed by one researcher, whereas this paper proposes a co-design process to include the views, needs, and goals of the target audience in the design of the EER. Finally, the work by Aragon [36] briefly discusses the underlying mechanisms of the EER, namely (1) the potentially positive effects of collaboratively solving nursing challenges and (2) the strong relatedness of the puzzles with the curriculum thus providing the students with alternative means to consolidate what was learned during the program [36]. The EER in this paper makes a clear choice to define the underlying mechanisms in the SDT theory and the flow theory.

There are more EERs for eMH, such as the one proposed in the study by Petkari and Calvo [37], and the number of EERs is likely to increase over time as its potential is currently widely researched. Similar to the studies discussed earlier, Petkari and Calvo [37] focused on EERs as alternative or additional elements for existing curricula, targeted at soon-to-be professionals. In those contexts, the design of the EER is mostly in the hands of educators or researchers, while the professionals are only involved in evaluations [37].

In a previous paper, we described the careful considerations we took to aim for a successful design and subsequent deployment of an EER that is tailored toward experiential learning about (innovative) digital elements suitable for blended therapies [38]. This paper describes the execution of a preliminary version of the EER.

This study aims to (1) further pursue the ambitions of a co-designed EER for training mental health care professionals in eMH, (2) investigate whether an EER would be an acceptable solution for this purpose, and (3) determine whether the SDT theory and flow theory are recognized as effective mechanics based on the experience of professionals, thus making a contribution to the field of eMH and the emerging field of EERs in the context of health care.

Co-Designing an EER for eMH Therapy

On the basis of these encouraging arguments and inspiring examples of eMH EERs in pursuing the development of an escape room, we gauged practitioners' first reactions and expectations regarding this idea. This delivered mostly positive responses: mental health care professionals expect an escape room to be innovative, engaging, inviting, and much more reasonable than the traditional and familiar eMH training methods. Moreover, based on our user requirements analysis [39], we aimed to develop a physical escape room. One argument is that a primary obstacle mental health care professionals reported in the user requirement analysis is that they are unacquainted with the possibilities of eMH. On the basis of this, we assumed that a physical room, with a facilitator present to moderate and explain, if necessary, would yield better results. A physical escape room provides more opportunities for collaboration [40]. Moreover, we learned from the user requirements analysis that users preferred a learning context in which they met others and could discuss with colleagues. Finally, playing an escape room is not a stand-alone solution. It should be part of a holistic experience, including a clear introduction and a reflection afterward, where participants can transfer the value of their eMH experience in the escape room to their real-life work practice [39].

As we aimed to develop a serious game for training mental health care professionals in using eMH, it is important to classify an escape room according to the gameplay/purpose/scope model for serious games [41]. This allows a better understanding of an escape room as a serious game and whether it meets our aim of developing a game with educational content. According to the definition by Djaouti et al [41], which distinguishes game-based from play-based activities by noting that the latter lacks a clear goal, we can, we can classify the gameplay of our intended escape room as game-based. The goal of the escape room is to enhance knowledge of the use of eMH by mental health professionals and train them to use eMH tools in different situations. That immediately brings us to the purpose of the escape room, which, in the classification by Djaouti et al [41], would be training. The intention of the escape room is to improve the performance of mental health professionals in using eMH, including how to use certain techniques as well as how to use these techniques in a particular context. We could also argue that there is a data exchange component in our escape room, as we also aim to encourage the users to help each other solve the puzzles and learn from each other's experiences. The scope of the escape room is health care, with mental health

professionals as the main target audience. However, there is also the possibility that an escape room to train eMH skills is incorporated into educational content at vocational or higher professional education [41].

Objectives

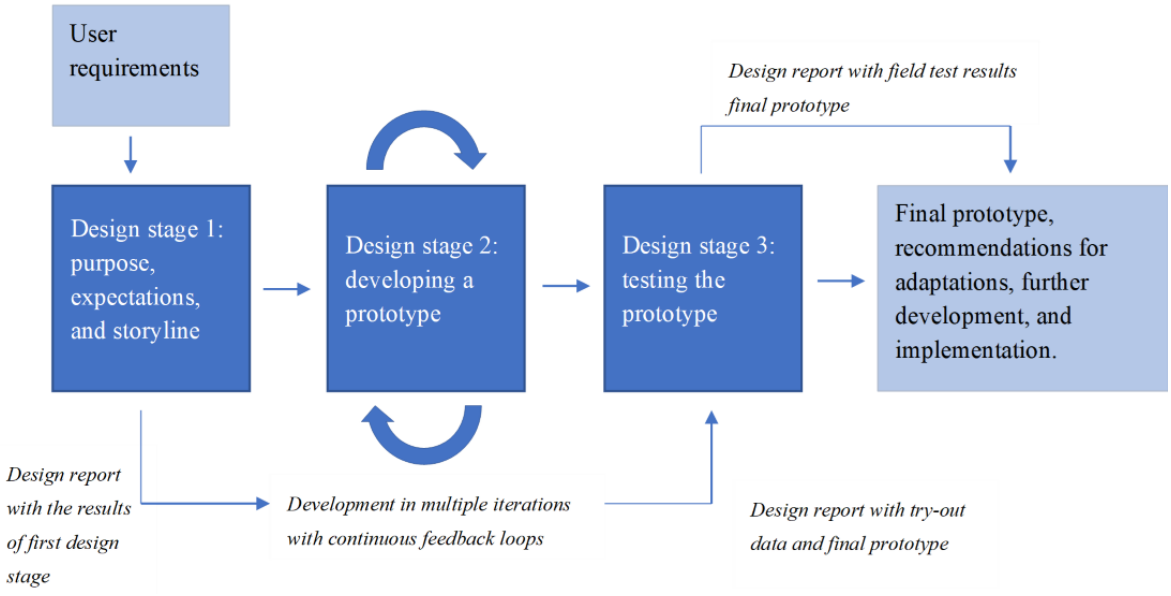
This paper presents the co-design process of developing an escape room for training mental health care professionals in eMH. By assessing the expectations and experiences regarding an EER in mental health care, we aim to evaluate the perceived usability and to gauge the effect of underlying mechanisms rooted in SDT and flow theory for training mental health care professionals' eMH skills.

Methods

Study Design

To shape the development and explore the perceived usability of an escape room, we used a process of 3 design stages, in which a testable prototype of an escape room for mental health care professionals was developed. In this "research through design" process, the design of the escape room is part of a research project in which the design activities contribute to generating knowledge [42]. Zimmerman et al [43] define research through design as "a research approach that employs methods and processes from design practice as a legitimate method of inquiry." This knowledge is applied at each stage to iteratively develop prototypes within the design process [42,43], leading to the final prototype. To explore the potential of an escape room, users' experiences and ideas are gathered on whether the escape room works as intended and whether it lives up to the promise of a concept for knowledge and skill enhancement. With this research through design approach, we seek to contribute to evidence-based decisions on the implementation, including potential adaptations and strategies to maximize the users' acceptance of the concept. This means that the design of an escape room is explored for suitability in the practice of mental health care and for its potential to address the aim of enhancing professionals' skills in using eMH. We organized 9 co-design sessions in 3 design stages (Figure 1). The escape room's purpose, expectations, and storyline ideas were formulated in the first design stage. These ideas were further translated into a gaming concept (ie, an escape room prototype), assessed for suitability in the second stage. In the third stage, a prototype was tested with end users (ie, mental health care professionals). Data from all stages within this process contributed to exploring the concept's perceived usability. In the Exploration of an Escape Room in 3 Design Stages subsection in the Results section, we report on the specific procedures and results of the co-design sessions. In doing this, we integrate the procedures and results to describe each design stage as a whole.

Figure 1. Three design stages in the development of an escape room for e-mental health (eMH) training.



Research Context

The study was conducted at a mental health care organization in Eindhoven (Geestelijke Gezondheidszorg Eindhoven en De Kempen, GGzE), in the southern part of the Netherlands. GGzE delivers specialized mental health care for people with various psychiatric and psychological issues (eg, depression, anxiety disorders, psychosis, autism, and personality disorders). Mental health care professionals working at GGzE include (specialized) psychologists, psychiatrists, nurses, and social workers. GGzE treats approximately 12,000 clients annually and has 2368 employees [44]. The research involved researchers from Tilburg University, Eindhoven Technical University, and Fontys University of Applied Sciences. Through this collaboration between universities, we combined expertise from social science, human and technology interaction, and game design. This combination of expertise vouches for a project in which all necessary aspects of a useful design process (eg, technique, human attitudes and behavior, and serious gaming concepts) and their interaction are considered.

Participants

Stages 1 and 2

The co-design sessions, including 1 web version in stage 2, were held with 5 mental health care professionals (psychologists: n=2, 40%; nurses: n=2, 40%; and social worker: n=1, 20%), 1 innovation staff member, 3 researchers, 3 research laboratory leaders, and 2 game designers. Together, they formed the co-design team for this stage in the design process (Table 1). In line with the existing literature on co-design, we involved stakeholders from the three areas of expertise that were needed for the design project [45-47]: (1) the mental health care professionals that represent the end users of the escape room, the innovation staff that will be involved in launching the escape room in practice; (2) the game designers who have expertise in the development of an escape room; and (3) researchers that are involved in exploring the usability of and experiences with an escape room. The mental health care professionals and the innovation staff members worked at GGzE. They were recruited through an announcement on the organization's internal communication platform and through network contacts of the project team, that is, a convenience sampling method was used. Mental health care professionals with varying levels of expertise were included in this stage of the development process. After their initial agreement, the participants were further informed about the project, the purpose of their participation, and how their data would be handled.

Table 1. Participants' roles in the development of an escape room for e-mental health (eMH) training.

Participants	Design stage 1	Design stage 2	Web version (design stage 2)	Design stage 3
Psychologists	• Participants (n=2)	• Participants (n=2)	—[a]	• Participant field study session (n=3)
Nurses and therapists	• Participant (n=1)	• Participants (n=2)	Participant (n=1)	• Participant field study session (n=2)
Social workers	• Participant (n=1)	—	—	• Participant field study session (n=5)
Mental health professionals' team for web-based treatment	—	—	—	• Participant test session 2 (n=4)
Innovation staff members	• Participant (n=1)	• Participants (n=2)	Participant (n=1)	• Participant test session 1 (n=2) • Facilitator field study sessions 1 to 3 (n=1)
Researchers	• Participants (n=3)	• Participant (n=1) • Observers (n=2)	Participant (n=1)	• Observant field study sessions 1-3 (n=2)
Research laboratory leaders	• Facilitator (n=1) • Participant (n=1) • Observer (n=1)	• Observers (n=3)	Participant (n=1)	—
Game designers	• Facilitator (n=1) • Participant (n=1)	• Facilitator (n=1) • Observant (n=1)	Facilitator (n=1)	—

[a]Not applicable.

Stage 3

In the third stage of the design project, we first asked a purposively sampled selection of mental health care professionals, namely 4 highly experienced eMH therapists from the team for web-based treatment and 2 members of the innovation staff of GGzE, to participate in 2 test sessions. The 2 test sessions secured a sufficient perceived flow of the escape room workshop (ie, including introduction and reflection). They eliminated further technical and practical issues concerning the logistics escape room (Table 1). Subsequently, we invited therapists from all disciplines within ambulatory specialized mental health care to participate in 3 field study sessions to test whether the final prototype of the escape room would be suitable as a concept to enhance mental health care professionals' skills in using eMH (Table 1). The first session included 2 psychologists, 1 nurse specialist, and 1 social worker. The second session involved 5 therapists engaged in social work and supporting clients' daily activities. Finally, a clinical psychologist and a verbal therapist tested the escape room in the third session. All mental health care professionals in the 3 field study sessions had some essential experience with eMH (ie, using a web-based module and video calling on a nonfrequent basis).

Data Collection and Analysis

Detailed notes were taken for all co-design sessions in all stages and were used to compose written reports. These reports were checked with the participants, after which themes were identified to describe, organize, and interpret the data [48]. In addition, video recordings were made of the sessions in stage 3, which were used to verify and complement the written reports. In the

first stage, the 2 sessions were analyzed individually because of the difference in nature of the sessions. In the second stage, the written reports were summarized into recommendations for the design team, which they used in the next iteration of the design. In the third stage, the first 2 test sessions were analyzed together, and the 3 field study sessions were also combined for analysis.

We used the Standards for Quality Improvement Reporting Excellence reporting guidelines to check the completeness and improve the quality of our manuscript [49]. A completed checklist can be found in Multimedia Appendix 1.

Ethical Considerations

This project was approved by the ethics review board of Tilburg University (EC-2013.14) and by the scientific committee of mental health care in Eindhoven (the Netherlands), where the participants were recruited. All participants were informed about the research and allowed to withdraw at any moment. All participants signed the informed consent form. Data in this project were anonymized and cannot be linked back to participants. Participation was voluntary, without financial compensation.

Results

Exploration of an Escape Room in 3 Design Stages

We describe the 3 stages of our research through the design process subsequently (Figure 1). As explained in the Methods section, we elaborate on the specific procedures we followed in each design stage to gather the data, the results generated, and how the findings within all 3 stages provide insights into the usability of an escape room as a game-based solution for eMH training for mental health care professionals (Figure 1).

Design Stage 1: Purpose, Expectations, and Storyline

Overview

In the first co-design session of design stage 1, which took place in December 2019, a joint purpose and common ground for the escape room were formulated. Participants were invited to work in subgroups to discuss the following: (1) the current situation, including the mental health care professionals' attitude regarding eMH; (2) a vision of how future mental health care professionals should be equipped and accommodated to integrate eMH into their daily practice; and (3) the requirements regarding an escape room to transition from the current to the future mental health care professional, in which eMH plays a central role, including the success and failure factors for this concept.

This was done by exchanging ideas, associating concepts, and explaining the relationship between eMH and the work of mental health care professionals.

Following this, a second co-design session was organized, aimed to deliver the storyline for the envisioned escape room. Information was gathered on the particulars of clients in mental health care, the mental health care setting in general, the various treatments and services available, including eMH, and the different client journeys that occur. This provided information on the requirements regarding the content of the game. The designers delineated a fictitious client and a story to form the basis of the escape room game. This initial storyline was verified with the mental health care professionals of the co-design team and was adjusted accordingly.

Purpose

In the first co-design session, the participants mentioned that many mental health care professionals are still convinced that the best way to treat a client is through face-to-face interaction. There is still a feeling of eMH being imposed by financial incentives, offering very limited job enrichment for mental health care professionals and very limited options to enrich treatment possibilities for clients. As a result, practitioners often need more interest or time to explore the possibilities of eMH. An important driver for mental health care practitioners to use eMH in the future would be that they see a clear-cut advantage in it for their clients, such as feeling less of a "patient" and gaining self-control and autonomy. This could be established by gaining positive experiences using eMH, for example, in training. Another aspect of the current attitude of many mental health care professionals mentioned in the co-design session was the lack of knowledge and skills to use eMH. Enhancing this knowledge requires time that needs to be facilitated by the organization they work for.

Despite this attitude, mental health care practitioners see the need to integrate eMH into therapy. The future mental health care professional is seen as someone for whom eMH is a natural part of the job. The session revealed that mental health care professionals consider their profession a craft, and some may see innovation as detrimental to this craft. According to the participants in the session, an escape room should demonstrate the opposite: instead of devaluating the craft, eMH has to reinforce it.

Expectations

On the basis of the purpose of the escape room, it is essential that sharing positive experiences and building confidence in using eMH are emphasized when designing a prototype. In the co-design session, it was stated that examples of positive experiences with eMH in clinical practice should be visible in the escape room, along with opportunities for professionals to share their positive experiences. Furthermore, the educational content of the escape room should seamlessly fit various real-life scenarios, making a transfer to everyday practice easy. Specific and well-defined learning goals should provide the opportunity to gain confidence in using technology, and the escape room should consist of customizable elements to fit the particular learning needs. Finally, the escape room should be engaging and contain continuous feedback loops, a recognizable storyline, and multiplayer options for a shared learning experience.

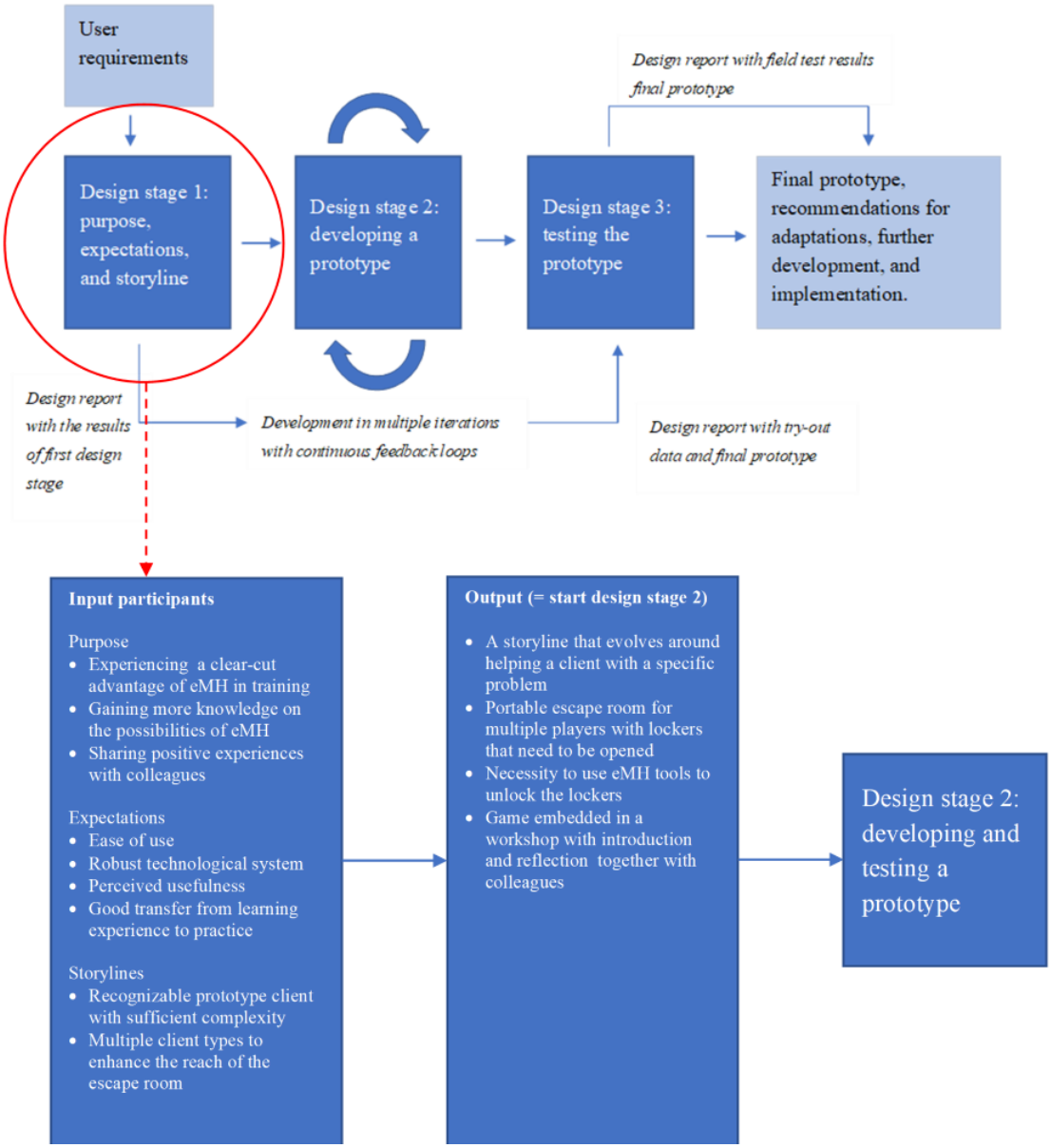
In addition to game-specific criteria, conditional requirements, such as perceived ease of use, a robust (technological) system, and perceived usefulness were considered significant. Furthermore, the social influence of colleagues, the perception of available time, endorsement by management, and an excellent incentive to engage in the escape room are issues that need to be considered. Finally, the escape room should be a complete learning experience with a clear introduction and reflection after

playing the game. Mental health care professionals must transfer from the learning experience to their therapeutic practice.

Storylines

On the basis of the input in the first session, the designers crafted a storyline for the escape room, which was presented to the co-design team. In their reflection on the storyline, the co-design team remarked that the mental health problems of the fictitious client should be more complex to be eligible for specialized mental health care. It should also contain more details to be sufficiently recognizable for therapists. In this stage of the design process, the co-design team also concluded that it would be valuable if there were sets of multiple storylines revolving around different settings or problems in mental health care to enhance the reach among the potential users of the escape room (Figure 2).

Figure 2. Results of the first design iteration in the development of an escape room for e-mental health (eMH) training.



Design Stage 2: Developing a Prototype

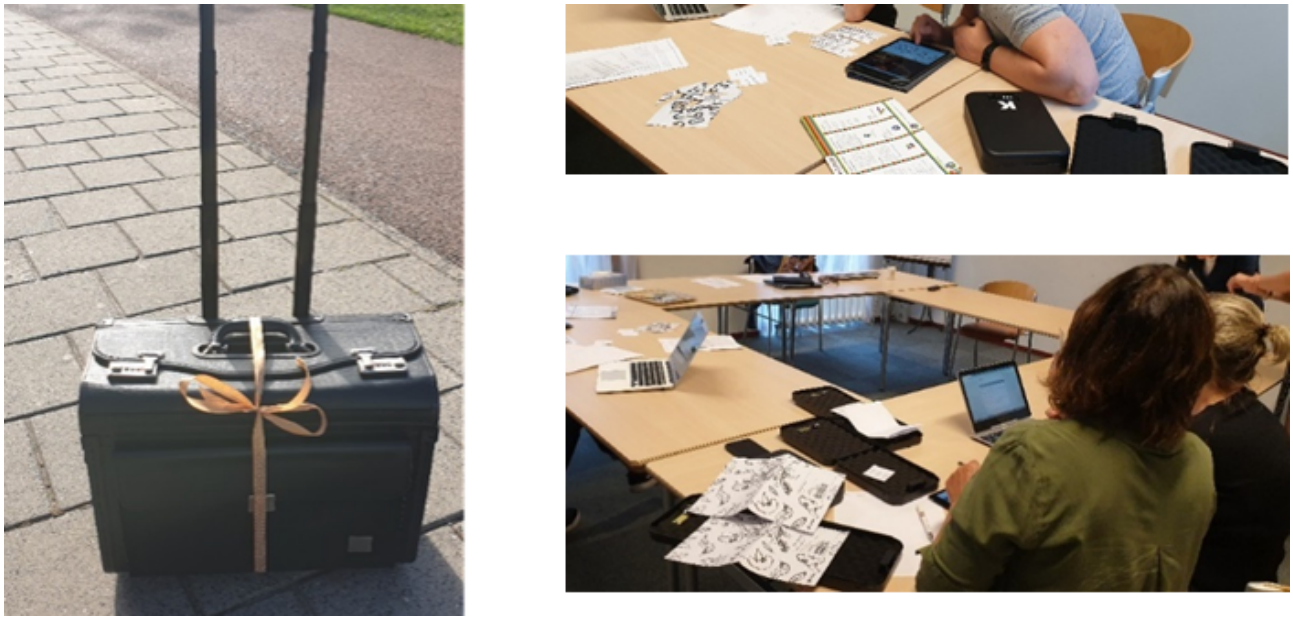
Overview

The prototype was presented to the co-design team in February 2020. The prototype of the escape room consists of 2 storylines, which are played separately. In each storyline, the participants are introduced to a client who experiences mental health problems. The challenge in the escape room is to answer questions in the client's electronic patient file, leading to a final solution to help the client take the next step. These questions can only be answered by finding clues in several locked safes, thoroughly analyzing the information given about the client, and using the eMH tools they see in the room. The eMH tools the participants provided to "help their client" were an electronic diary, a VR application for exposure therapy, biofeedback, a chat function, and web-based treatment platform. All materials used in the escape room are portable, which means playing the escape room is not dependent on a fixed place (Figure 3). The actual game is embedded in a workshop with an introduction to familiarize with the topic, goals, and expectations regarding

the escape room and a reflection afterward to discuss the experiences and take-home messages.

In a try-out session, the co-design team tested this prototype by playing the escape room in 3 small groups of 3 participants. This delivered additional data for our research regarding the expectations and experiences of an escape room. At the same time, it provided the designers with feedback for a second prototype. In each round, the participants were given 15 minutes to play the game and solve the puzzles. Each subgroup reflected individually on the prototype directly after playing the escape room. At the end of the session, a plenary discussion took place to summarize the results. The designers developed a second prototype for the escape room based on the feedback. In this phase, a second storyline was also developed. Different eMH tools were added and various connections were made between these eMH tools, the storyline, and the puzzles.

Figure 3. Picture of an escape room setting.



In March 2020, the world was confronted with an unprecedented global pandemic that caused a significant disruption of life as we knew it. Burdened by the consequences of the pandemic and the measures that national governments took to control the virus (eg, social distancing, lockdown, and quarantine), the activities in our design process were also heavily disturbed. Switching to this new situation, we sought to continue the design process by developing a web-based alternative to test the second prototype of the escape room in April 2020. This web version was a digital copy of the physical room, containing the same elements, puzzles, and eMH tools, except for the VR goggles. This second prototype included 2 storylines, each with a fictitious client, and was tested by the co-design team members. One storyline was tested while being moderated by the designers, the other storyline was tested by each member individually, without being moderated.

Try-Out Sessions

The feedback gathered during the try-out sessions delivered several expectations and experiences of mental health care professionals regarding an escape room. First, storylines in an escape room benefit from a good match with a client's pathway in mental health care and reflect truly relatable situations. Moreover, it is essential to focus on mental health problems as an integrated part of everyday life situations. This is important because otherwise, the suggested eMH tools cannot be appropriately linked, and a genuine eMH experience will be missed. This suggestion was also given in response to the final storyline in the previous phase of the design process. Second, there should not be too much focus on solving the puzzles in the escape room, as it diminishes the awareness of eMH in the game. More importantly, the eMH tools should be linked explicitly to the puzzles. The participants should experience a (positive) consequence when choosing, for example, finding the next clue when trying a specific eMH tool or experiencing that someone (ie, the fictitious client) benefits from you as a practitioner choosing to use eMH. Third, variation and differentiation in an escape room vouches for a stronger and fuller user experience. This could be established by multiple storylines (as was stated in the previous phase of the design process), different sets of eMH tools that can be used, and variations in the level of difficulty in the escape room. In this way, the escape room can be played multiple times and addresses different levels of expertise, either regarding eMH or playing an escape room. Fourth, there should be a clear conclusion to the escape room and, consequently, a euphoric feeling when finishing the game. To this end, the purpose at the beginning of the game may need to be more apparent. Fifth, it should be beyond any doubt that the experience of eMH in the escape room can also be transferred to the daily practice of a mental health care professional. Otherwise, the escape room will not be considered credible.

Web Version

From the tryout, it became clear that the experience of a web version of an escape room does not closely resemble the experience of an escape room in real life. Due to several technical hassles, the game's flow was hindered multiple times.

Most importantly, it impeded the possibility of team interactivity, working together, and exchanging ideas about eMH, which was mentioned multiple times in the design process as one of the critical success factors for an escape room.

Final Prototype

After receiving the feedback on the web version of the escape room, the designers continued developing a prototype that was fit for testing by end users. The co-design team subsequently assessed a final prototype in July 2020. After minor adaptations, this prototype was finalized in September 2020. Following this, a field study was planned (stage 3) to gather the user experiences of the escape room among the end users, that is, mental health care professionals from various disciplines (Multimedia Appendix 2).

Design Stage 3: Testing the Final Prototype

Test Sessions

In 2 test sessions (September and October 2020), we first checked the flow of the escape room workshop and the functioning of the materials and technology. This delivered a positive outcome regarding the potential of the concept for exploring eMH and, in that way, enhancing mental health care professionals' knowledge about the possibilities of technology in therapy. The participants of these sessions shared several thoughts on how to improve the workshop before starting the field study sessions with end users.

The first finding in the test sessions was the importance of good functioning of the content and technical aspects of the escape room to ensure a good flow in the game. Besides a good flow in the game, a facilitator should lead a good introduction and reflection to complement the learning experience.

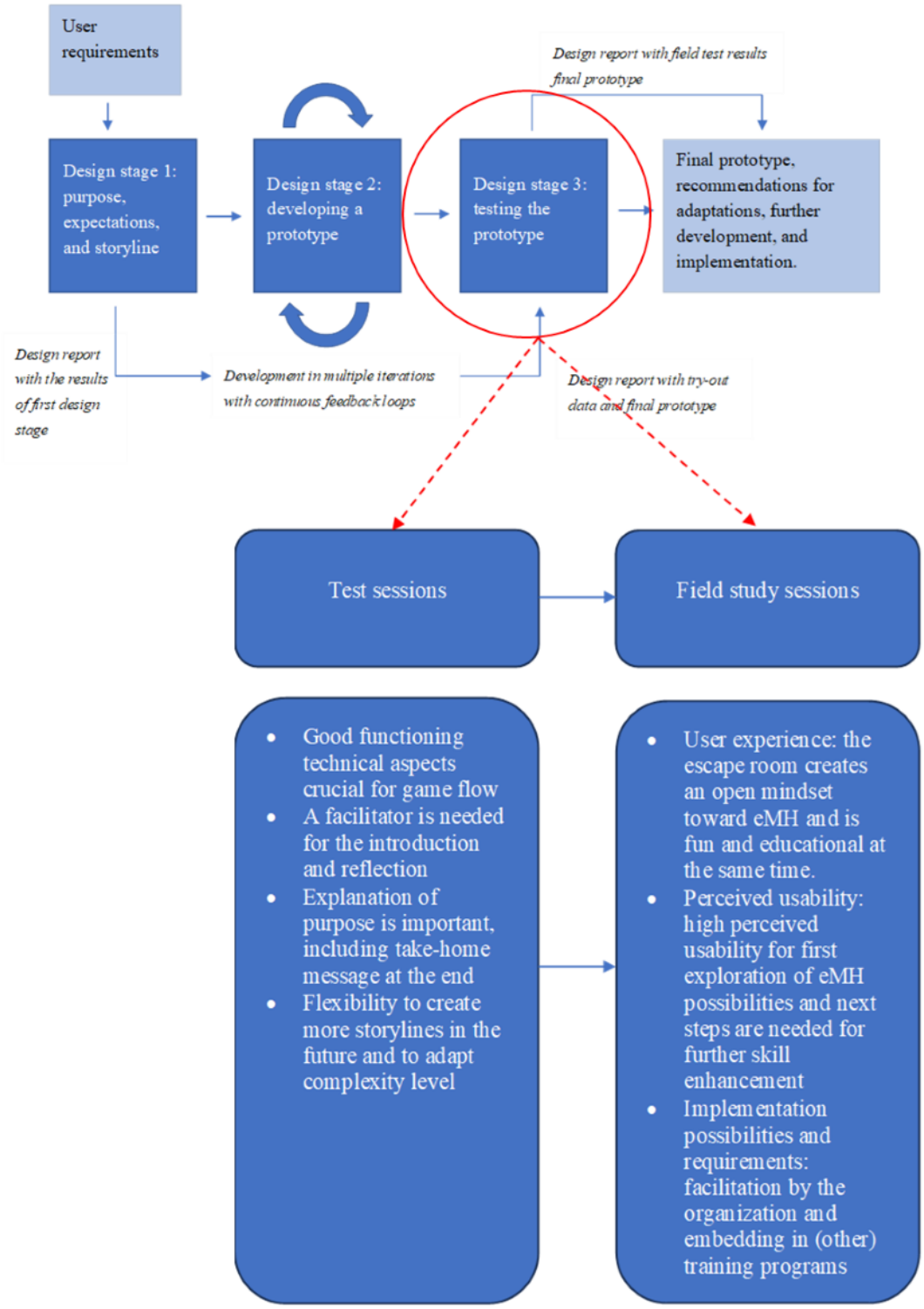
According to the participants in the test sessions, a good explanation of the purpose and relevance of the escape room and presenting a game manual to maximize the effect of playing the escape room is necessary. One participant particularly mentioned the importance of stressing the educational purpose and preventing the feeling that the workshop is merely something for fun. After playing the escape room, reflecting on the purpose is important. According to the participants, there should be a clear take-home message, discussion on the experiences gained in the escape room, and acquired knowledge to share with colleagues afterward.

Another issue raised during the 2 test sessions was the need to continuously adapt the storylines or create the flexibility to insert additional storylines in the escape room game. Two storylines may be too few to address all therapeutic contexts, all levels of complexity in mental health care, and all eMH possibilities that could add value. The participants also marked a sense of reality in the storylines. Opportunism should be avoided because it should not imply that technology will always vouch for a positive treatment outcome. For example, in VR, the escape room should not pretend that it is something that instantly makes everything better. This was regarded as an essential issue to warrant the credibility of the escape room. However, the same session also mentioned that experiencing eMH tools in this manner created awareness about escape room's potential. At the same time, the biofeedback application also showed that it is possible to visually track panic attacks in such an application. The final aspect regarding the mentioned storylines is using the correct professional jargon in the escape room. At one point, a participant stated that the way things were written in the escape room did not reflect the language a therapist would use. In contrast, language and jargon appeared very personal, making it difficult to always be accurate in an escape room. This emphasizes the fact that there is a large diversity among mental health care professionals and the clients they are treating.

Finally, the participants discussed the complexity of the game itself. Some participants found figuring out the puzzles and clues challenging, but others remarked it was too simple. Some solutions were easily found without using the eMH tools in the escape room, resulting in an ineffective learning experience. Some participants in the test sessions suggested creating multiple levels of complexity in the escape room to address several types of game players (ie, level of experience with escape room games) and other kinds of therapists.

On the basis of the results of these 2 test sessions, minor adjustments were made regarding the complexity of the puzzles to ensure that clues and solutions could not be found without using eMH tools. Furthermore, the introduction and reflection were refined to embed the escape room game in a workshop format (Figure 4).

Figure 4. Results of the test and field study sessions (third iteration) in the development of an escape room for e-mental health (eMH) training.



Field Study Sessions

Overview

Hereafter, the workshop was tested in 3 sessions with mental health care professionals from different areas of expertise. A total of 11 mental health care professionals participated in these field study sessions (November and December 2020 and June 2021) and were asked to fill in a questionnaire (Multimedia Appendix 3) at the end of the workshop. Their answers formed the basis for a discussion about their user experiences. The questionnaire consisted of an overall experience rating (question 1), three parts to reflect on (1) the quality of the introduction (questions 2-4), (2) the game itself (questions 5-11), and (3) the reflection (questions 12-15), overall questions regarding the perceived usability (questions 16-19), and a question on strengths and suggestions (question 20). The questions in the 3 parts of the questionnaire were constructed based on the results of the co-design sessions in design stage 1. As we explained in design stage 1, mental health care professionals use eMH mainly because of its apparent benefits. To encourage this, the escape

room should therefore highlight positive experiences with eMH in clinical practice, in line with the flow theory that states that a positive learning experience can be reached if the challenges meet the abilities of the users and if they can share this with colleagues [31]. To explore the extent to which the escape room was experienced positively, we formulated questions 1, 3, 9, 10, and 13. Furthermore, from design stage 1 and SDT, we learned that the educational content must align with real-life scenarios, making it easy to apply in practice [29]. Questions 6, 8, and 15 were therefore constructed to assess whether the storylines are recognizable and whether the training with eMH tools in these storylines connects to their daily practice. In addition, questions 2 and 5 were formulated to explore if the learning goals become clear, and questions 7 and 12 assessed whether the escape room helps to build confidence in using technology, with customizable elements tailored to specific needs. Finally, the game should enable multiplayer options for shared learning [31]. To assess whether the escape room was experienced that way, we added questions 4, 11, and 14. Together this led to the questionnaire, which had 20 targeted questions. The answers to the questions were subsequently used as a guide in discussion with the participants to gain a more in-depth understanding of the possibilities of an escape room as an environment for mental health care professionals to experiment and practice with eMH.

User Experiences

From the questionnaire and discussion, we learned that the participants in the field study sessions were enthusiastic about the workshop they participated in. They stated that the workshop is inspiring and creates a positive mindset toward eMH. They recognized the potential to remove barriers they had at first, showing that eMH is accessible and not always that difficult to use. The participants experienced significant fun in the escape room, making it feel less obligatory. This is expected to help people engage more easily. In addition, the alternation between playing and reflecting was mentioned as meaningful. In total, 2 (18%) of the 11 participants mentioned they were pleased that there was no explicit element of competition in the escape room. According to these participants, too much competition, for example, ranking who has solved the escape room the fastest, would distract the users from the storyline and the eMH possibilities they come across.

All but one participant expressed that they experienced the storylines in the escape room as realistic. One mental health care professional did not recognize the storylines in the escape room. Another aspect related to this was the “credibility” of the suggested effect of eMH in the escape room. Although most (9/11, 82%) of the participants could describe the use of eMH to certain benefits for mental health care, it should be carefully suggested in the escape room. However, most (9/11, 82%) participants experienced how they could practically apply eMH in therapeutic situations. For example, VR was experienced in the escape room, resulting in a discussion on how VR could be used for clients in specific treatment situations. The participants were able to connect their VR experience in the workshop to real-life situations and could see the added value of using eHealth. All kinds of eHealth suggestions were exchanged, which illustrates the effect of being given the possibility to explore and experiment with eHealth in this type of playful setting.

Participants wondered whether an escape room would appeal enough to therapists who were skeptical about eMH. The participants also mentioned that the current storylines and eMH tools may not be recognizable for all mental health care professionals. It was considered important to explore eMH tools that are also accessible in real life to be able to transfer knowledge from training to reality. In contrast, it is also mentioned that mental health care professionals with vast experience in eMH will find it more attractive to explore more experimental eMH tools.

Perceived Usability of an Escape Room for eMH Training

To explore the perceived usability of an escape room for skill enhancement in eMH, we asked the participants to indicate whether the workshop provided them with a more elaborate and precise picture of the possibilities of eMH (Table 2). The feedback on this issue varied among the participants. In total, 2 (18%) of the 11 mental health care professionals reported that they had not gained a better picture of the options of eMH in the escape room. This was because they already had a relatively good basic knowledge about eMH beforehand. The participants who were less experienced users of eMH reflected that they had a much better idea about the possibilities of eMH after participating in the escape room workshop. It was suggested that the current escape room prototype would benefit from further development to create more options to try additional eMH tools because, in the current prototype, not all mental health care professionals would find tools suitable for their therapeutic practice.

Table 2. Perceived usability of an escape room (ER) for e-mental health (eMH) training (questionnaire results).

Items	Score on a scale of 1-5, mean (SD)
Suitability of an ER to explore possibilities of eMH	4.7 (0.92)
Possibilities for collaboration in the ER	4.7 (0.47)
Clarity of the storylines	4.3 (0.92)
Clarity of learning goal in the ER	4.6 (0.67)
Balance educational content and fun	4.7 (0.68)
Usefulness for daily practice	3.8 (0.77)

The participants largely agreed on the suitability of an escape room concept for exploring eMH (average score of 4.7, SD 0.92). They perceived the learning goal in the workshop as evident (average score of 4.6, SD 0.67). They also experienced a good balance between the educational and the fun part in the escape room (average score of 4.7, SD 0.68). It was mentioned in the discussion that suitability is supported when the bridge between the escape room and everyday practice is explicitly discussed in the workshop as well as the steps that can be taken after completing the workshop and returning to the workplace. Follow-up training is essential because otherwise the acquired knowledge will not be secured. According to the participants, this escape room offers an excellent first experience with the possibilities of eMH; more than that, it is just another educational workshop, "It is more 'getting to know eMH' than 'learning how to apply eMH." It was also suggested that more focus on the client groups that people are treating would be helpful in the reflection part of the workshop. The participants advocated that involving professionals with similar client groups or treatment contexts would be more effective in stimulating the exchange of experiences with eMH and the transferability of the experiences in the escape room to daily practice.

In line with the findings about follow-up training, the participants elaborately reflect on how to apply and secure knowledge on eMH once the first experiences are gained in the escape room. While this workshop helps to talk about eMH, keeping the conversation going in everyday practice is equally important. According to the participants in our study, mental health care professionals are focused on the content of their treatment and support, and the knowledge gained in the eMH escape room might lose its impact. It was stated that the organization should provide continuous information about what is new in the field of eMH and should provide time to invest in exploring new tools. The concept of an escape room or a game offers a very persuasive opportunity to invest in keeping knowledge up to date and motivating people. All workshops suggested that this would be an excellent fit for a teambuilding activity.

Implementation Possibilities and Requirements

In addition to gathering experiences with the escape room workshop, we asked the mental health care professionals who participated to indicate what they deemed essential for successfully implementing it. The first thing mentioned was that professionals in mental health care should be confronted with eMH from various directions, that is, in an escape room, but also by clients, colleagues, and organizational policies aimed at being innovative, requiring innovative and digitally skilled employees. Incorporating an escape room in an introductory meeting might be possible when people are new to the organization.

Another issue that emerged from the workshops concerns the approachability of the escape room. It is experienced as convenient that the escape room is currently a mobile version, which does not necessarily require people to go somewhere and spend much time traveling. It would be even more convenient if the escape room could be played without an external facilitator so that participants could be more flexible in their time planning. However, this remains the main concern of the participants in the workshops: will mental health care professionals be provided with enough time to explore eMH in the future, and will their production targets be adjusted accordingly? In addition, it should be decided who will provide training services and materials in the organization. Finally, the participants stated that a strong vision by the policy makers on the use of eMH in the organization and the expectations regarding the future mental healthcare professional is indispensable to warrant a successful implementation of an escape room or any serious gaming concept for eMH training in mental health care.

Discussion

Principal Findings

In this research through design study, we investigated the perceived usability of an escape room for training mental health care professionals in using eMH. On the basis of the results in each stage of our research and the overall findings, we can conclude that an escape room concept might be suitable for this type of training. We subsequently discuss our findings by reflecting on the experiences with our escape room prototype and what this means in light of the purpose of our design project: developing new training methods for mental health care professionals, with the ultimate goal of contributing to increasing the adoption of eMH. We conclude with an overall view of the possibilities of game-based training in mental health care based on the results of this design study.

What we know from earlier research, which also emerged in this study, is that many mental health care professionals are not convinced of the benefits of eMH, resulting in dismissive attitudes and behavior [50,51]. The escape room was experienced as both fun and educational; however, most importantly, it induced a more positive stance toward eMH tools. While an open mindset is a prerequisite for change, the escape room may offer a potential concept as the first step to changing attitudes and behavior among mental health care professionals regarding eMH [36,40].

For the escape room to be accepted as a valuable training method by its end users, the storylines, game elements (puzzles and clues), and eMH tools in the escape room needed to be credible and recognizable for mental health care professionals. At the same time, it became clear that it is tough to design storylines that seamlessly match the practices of mental health care therapists. In one of the first versions of the prototype, there was too much focus on everyday life situations instead of mental health problems. This is crucial because the participants need a genuine eMH experience to link to their treatment context properly. This feedback was given in different iterations in the design process, which marks the difficulty of designing a good and representative storyline and the importance of involving multiple disciplines in a co-design team [46,47].

Moreover, it became clear that mental health care professionals are diverse in terms of their client population, type of mental health care services, and level of adoption in eMH. This requires a flexible and adaptable training concept to expand the reach of the escape room. Greater diversification regarding the

storylines and available eMH tools in the escape room may be needed, along with various difficulty levels. The participants in the field study sessions indicated that the escape room in its current form might not be suitable for people who are already frequent users of eMH. This implies that the gains of the current escape room are probably highest for people who are relatively inexperienced eMH users, which applied to most mental health care professionals when we started the design process. Adding new, more innovative tools could, however, expand the reach of the escape room.

The participants welcomed the escape room as an engaging concept and noted that its design needs continuation and further development. In its current form, the escape room offers an environment for exploration and the first acquaintance with eMH rather than actual skill development in applying these tools. We believe that, whereas most mental health care professionals are—partly because of the coronavirus pandemic—familiar with standard eMH tools that are also used in everyday life (eg, videoconferencing and SMS text messaging), they remain unaware of the possibilities of more novel technologies, such as VR and biofeedback, and may be biased about its use in their daily clinical practice. Therefore, creating a different mindset remains an important foundation for further skill development in the future and increasing adoption in general. An extensive research base reports on various barriers that prevent accelerating the adoption of eMH [5,51,52]. This escape room addresses one of the critical barriers, a perceived lack of knowledge. However, issues such as compensating for the loss of nonverbal information, the lack of trust in technological systems, and the availability of time and resources to gain positive experiences in using eMH also need to be addressed.

When exploring ideas for further training for specific eMH tools, the question is whether an escape room would also be suitable for this particular type of training. Looking at escape rooms in health care, we find that escape rooms are most appropriate for exploration and awareness and less ideal for training-specific skills [23,24]. This was also why we chose an escape room concept in the first place; it suited the results of our user requirements analysis that revealed a specific need for exploring and enhancing knowledge rather than specific hands-on skills. Training these hands-on skills should be addressed for the next step, particularly because the participants in our study also indicated this need. Moreover, for both the possibility of exploration and specific skill training, the support in terms of time and resources provided by the organization and its management is crucial. The organization's commitment to a digital transition and the necessity for mental health care professionals to be equipped with 21st-century skills must be at the forefront. Finally, follow-up research is required to test the escape room's effectiveness regarding knowledge increase and skill enhancement.

The Perceived Usability of Serious Gaming for eMH Training in Mental Health Care Based on Users' Experiences

The results of our study show positive experiences with an escape room as a serious gaming concept for eMH training in mental health care. These positive experiences might be explained in terms of the SDT [29] and the flow theory [31]. By attempting to address the needs of competence and autonomy, the users may have had a positive experience. The flow of the escape room may have positively contributed to their motivation to enhance their knowledge of eMH and their general attitude toward using eMH in their clinical practice [30].

In the design of the escape room, we addressed several important game mechanics that we found to have influenced our users' experiences, for example, an adequate level of complexity in the game [19,53,54]. The participants in our study emphasized that when a puzzle in the escape room takes too much time or is too difficult, it may lead to frustration and an adverse effect. The opposite of this also works negatively; when clues are too obvious, people may not feel they are taken seriously. In general, our escape room was experienced as having an appropriate level of difficulty, which led to a positive experience. We also paid much attention to the importance of the story or narrative in the escape room. When this narrative is too farfetched or not relatable to clinical practice, it loses strength; when people can identify with the story in the game, it may be much more effective than any other learning method [29-31]. In addition, the aspect of social interaction [30] was predominantly present in our escape room and experienced as one of the critical factors for its success.

Strengths and Limitations

In this study, we particularly choose a research through design approach to develop and test an escape room as a serious gaming solution for eMH training among mental health care professionals. The strength of this approach is that it enabled us to use formative evaluations [55] between the stages of design to provide input for the next design stage. We strongly believe that without these intermediate assessments, we could not design a prototype suitable to the practice of mental health care providers. More importantly, the involvement of mental health care professionals in our co-design team ensured a prototype that matched the therapeutic practice we pursued to reflect in our escape room. However, there are some limitations. First of all, we used an inductive analysis method, which allowed us to gather any information deemed necessary for the development of the escape room but may have lacked the required structure to establish an evaluation of the effectiveness of the escape room, in terms of skill enhancement, and for other researchers to replicate the analyses. Analysis procedures that would allow an assessment of the escape room's success could be to use a pre- and posttest quasi-experimental design for the evaluation [56], a controlled experiment with naturalistic and summative evaluations to assess whether the purpose of the game is indeed reached [57] or a heuristic evaluation using predefined rules of thumb to measure the usability of the escape room [58]. A second limitation of this study is that we included a small sample of mental health care professionals who all worked at the same mental health care organization. In other words, we did not include participants from a diverse mental health care setting other than the differentiation in mental health care contexts within an organization. In addition, we are aware that for projects like this, where participants are recruited through open invitation, people who are already enthusiastic about the topic,

in this case, eMH, tend to be the ones who sign up for participation. In future research, it would be worthwhile to test the escape room or any other game-based training solution among mental health care professionals who are skeptical about eMH to see whether also, in this group, the effect of creating an open mindset through exploration and experience-based learning is achieved.

Conclusions

In general, we can conclude that experience-based training, such as game-based training in an escape room, helps people experience what they are being trained to do or use. This experience may be much stronger than the experience of learning through traditional learning (eg, classroom, textbook, or e-learning). When training is more fun and engaging, this will increase the learning experience. Moreover, it increases the participants' motivation and creates an open mindset. In our study, we learned that an escape room as a serious gaming concept stimulates the willingness to participate in training and be open to change. This openness to change is essential in increasing the use of eMH. However, it is important that there is a clear next step to this first exploration, providing professionals with the opportunity to thoroughly train their skills in specific eMH tools. Subsequently, it would be valuable to evaluate the effectiveness of game-based training in an escape room for enhancing particular skills. Furthermore, designing is a cooperation between different experts. The design process should be viewed from different angles and areas of expertise. The contribution of mental health care professionals was necessary to create a credible game in which both the storyline and set of integrated eMH tools can be transferred to their therapeutic practice. Finally, the success of a game-based training method, such as an escape room, depends on the commitment of the organization in which it is implemented. People need to feel that they have the time and support of their management to engage in training. This commitment is more than just telling them to "go and participate in training." The whole organization must enhance specific knowledge based on common ground and a shared vision of what is needed regarding competencies and skills by mental health care professionals.

Acknowledgments

The authors would like to thank all mental health care professionals and other participants who participated in this project. This work is supported by the Netherlands Organisation for Scientific Research grant 055.16.141, under the project titled Psychosocial Games Supporting Mental Health Professionals in Stress Reduction and Empathic Interactions in Remote/online Psychotherapy, under the Serious Games for Professional Skills Program.

Data Availability

The datasets generated and analyzed during this study are not publicly available due to agreements about confidentiality and exchange of data that were made with the respondents in the study. The consent given by the participants does not allow for public availability. The data can be made available from the corresponding author on reasonable request and with the consent of the participants in this study.

Authors' Contributions

JB, MAF, and RJWST organized and conducted the co-design sessions. JB coded and analyzed the data and composed an initial draft for the manuscript. All authors contributed to multiple revisions. All authors agreed with the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
SQUIRE (Standards for Quality Improvement Reporting Excellence) checklist.
[DOCX File , 34 KB - formative_v9i1e58650_app1.docx]

Multimedia Appendix 2
Results of the second design iteration in the development of an escape room for e-Mental health training.
[DOCX File , 23 KB - formative_v9i1e58650_app2.docx]

Multimedia Appendix 3
Evaluation questions (English translation).
[PDF File (Adobe PDF File), 164 KB - formative_v9i1e58650_app3.pdf]

References

1. 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 01;13(3):288-295 [FREE Full text] [doi: 10.1002/wps.20151] [Medline: 25273302]
2. de Grood C, Raissi A, Kwon Y, Santana M. Adoption of e-health technology by physicians: a scoping review. J Multidisciplin Healthc 2016 Aug;Volume 9:335-344. [doi: 10.2147/jmdh.s103881]
3. Glueckauf RL, Maheu MM, Drude KP, Wells BA, Wang Y, Gustafson DJ, et al. Survey of psychologists' telebehavioral health practices: technology use, ethical issues, and training needs. Prof Psychol Res Pract 2018 Jun;49(3):205-219. [doi: 10.1037/pro0000188]
4. Lal S, Adair CE. E-mental health: a rapid review of the literature. Psychiatr Serv 2014 Jan 01;65(1):24-32. [doi: 10.1176/appi.ps.201300009] [Medline: 24081188]
5. Feijt MA, de Kort YA, Bongers IM, IJsselsteijn WA. Perceived drivers and barriers to the adoption of eMental health by psychologists: the construction of the levels of adoption of eMental health model. J Med Internet Res 2018 Apr 24;20(4):e153 [FREE Full text] [doi: 10.2196/jmir.9485] [Medline: 29691215]
6. Vis C, Kleiboer A, Prior R, Bønes E, Cavallo M, Clark SA, et al. Implementing and up-scaling evidence-based eMental health in Europe: the study protocol for the MasterMind project. Internet Interv 2015 Nov;2(4):399-409. [doi: 10.1016/j.invent.2015.10.002]
7. Wozney L, Newton AS, Gehring ND, Bennett K, Huguet A, Hartling L, et al. Implementation of eMental Health care: viewpoints from key informants from organizations and agencies with eHealth mandates. BMC Med Inform Decis Mak 2017 Jun 02;17(1):78 [FREE Full text] [doi: 10.1186/s12911-017-0474-9] [Medline: 28577543]
8. Kilbourne AM. E-health and the transformation of mental health care. Psychiatr Serv 2012 Nov;63(11):1059. [doi: 10.1176/appi.ps.631103] [Medline: 23117506]
9. 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]
10. Chan CV, Kaufman DR. A framework for characterizing eHealth literacy demands and barriers. J Med Internet Res 2011 Nov 17;13(4):e94 [FREE Full text] [doi: 10.2196/jmir.1750] [Medline: 22094891]
11. Norman C. eHealth literacy 2.0: problems and opportunities with an evolving concept. J Med Internet Res 2011 Dec 23;13(4):e125 [FREE Full text] [doi: 10.2196/jmir.2035] [Medline: 22193243]
12. Schalken F. Handboek Online Hulpverlening: Met Internet Zorg en Welzijn Verbeteren. Houten, The Netherlands: Bohn Stafleu Van Loghum B.V; 2013.
13. Bierbooms JJ, van Haaren M, IJsselsteijn WA, de Kort YA, Feijt M, Bongers IM. 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]
14. Crawford A, Gratzer D, Jovanovic M, Rodie D, Sockalingam S, Sunderji N, et al. Building eHealth and telepsychiatry capabilities: three educational reports across the learning continuum. Acad Psychiatry 2018 Dec 11;42(6):852-856. [doi: 10.1007/s40596-018-0989-0] [Medline: 30311178]
15. Echelard JF, Méthot F, Nguyen HA, Pomey MP. Medical student training in eHealth: scoping review. JMIR Med Educ 2020 Sep 11;6(2):e20027 [FREE Full text] [doi: 10.2196/20027] [Medline: 32915154]
16. Timman A, Caris JA. Leren in een andere realiteit. Onderwijs en Gezondheidszorg 2022;1:20-23. [doi: 10.24078/oeng.2022.2.128459]
17. Connolly TM, Boyle EA, MacArthur E, Hainey T, Boyle JM. A systematic literature review of empirical evidence on computer games and serious games. Comput Educ 2012 Sep;59(2):661-686. [doi: 10.1016/j.compedu.2012.03.004]
18. Squire K, Jenkins H. Harnessing the power of games in education. Insight 2003;3:5-33.
19. Wang R, DeMaria SJ, Goldberg A, Katz D. A systematic review of serious games in training health care professionals. Simul Healthc 2016 Feb;11(1):41-51. [doi: 10.1097/SIH.0000000000000118] [Medline: 26536340]
20. Susi T, Johannesson M, Backlund P. Serious games – an overview. Serious Games Cluster and Business Network Project. 2007 Feb 5. URL: https://www.diva-portal.org/smash/get/diva2:2416/FULLTEXT01.pdf [accessed 2024-12-16]
21. Makri A, Vlachopoulos D, Martina RA. Digital escape rooms as innovative pedagogical tools in education: a systematic literature review. Sustainability 2021 Apr 20;13(8):4587. [doi: 10.3390/su13084587]
22. Ritterfeld U, Cody M, Vorderer P. Serious Games: Mechanisms and Effects. Milton Park, UK: Taylor & Francis; 2009.
23. Adams V, Burger S, Crawford K, Setter R. Can you escape? Creating an escape room to facilitate active learning. J Nurses Prof Dev 2018;34(2):E1-E5. [doi: 10.1097/nnd.0000000000000433]
24. Mullen M, Seiler SO. Escaping the confines of traditional instruction: the use of the escape room in a transition to practice program. J Nurses Prof Dev 2019;35(6):354-356. [doi: 10.1097/nnd.0000000000000571]
25. Veldkamp A, van de Grint L, Knippels MC, van Joolingen WR. Escape education: a systematic review on escape rooms in education. Educ Res Rev 2020 Nov;31:100364. [doi: 10.1016/j.edurev.2020.100364]
26. Kolb DA, Boyatzis RE, Mainemelis C. Experiential learning theory: previous research and new directions. In: Sternberg RJ, Zhang LF, editors. Perspectives on Thinking, Learning, and Cognitive Styles. New York, NY: Routledge; 2014.

27. Prensky M. Fun, play and games: what makes games engaging. In: Digital Game-Based Learning. New York, NY: McGraw-Hill Education; 2001:5-31.
28. Grande-de-Prado M, García-Martín S, Baelo R, Abella-García V. Edu-escape rooms. Encyclopedia 2020 Dec 22;1(1):12-19. [doi: 10.3390/encyclopedia1010004]
29. Deci EL, Ryan RM. Handbook of Self-Determination Research. New York, NY: University of Rochester Press; 2004.
30. Vorderobermeier A, Abel J, Sailer M. Theoretical foundations and approaches in research on educational escape rooms: a systematic review. Educ Res Rev 2024 Aug;44:100625. [doi: 10.1016/j.edurev.2024.100625]
31. Csikszent M. Flow: The Psychology of Optimal Experience. New York, NY: HarperCollins; 1991.
32. Deen M. G.A.M.E., Games autonomy motivation & education: how autonomy-supportive game design may improve motivation to learn. Technische Universiteit Eindhoven. 2015. URL: https://pure.tue.nl/ws/files/3895781/784813.pdf [accessed 2024-12-07]
33. Okuda Y, Bryson EO, DeMaria SJ, Jacobson L, Quinones J, Shen B, et al. The utility of simulation in medical education: what is the evidence? Mt Sinai J Med 2009 Aug 29;76(4):330-343. [doi: 10.1002/msj.20127] [Medline: 19642147]
34. Reeves B, Read JL. Total Engagement: How Games and Virtual Worlds Are Changing the Way People Work and Businesses Compete. Brighton, MA: Harvard Business Review Press; 2009.
35. Rodriguez-Ferrer JM, Manzano-León A, Cangas AJ, Aguilar-Parra JM. A web-based escape room to raise awareness about severe mental illness among university students: randomized controlled trial. JMIR Serious Games 2022 May 05;10(2):e34222 [FREE Full text] [doi: 10.2196/34222] [Medline: 35511232]
36. Aragon CJ. Determining the effectiveness of an escape room in a mental health nursing course: a quantitative study. Northcentral University. 2023. URL: https://www.proquest.com/openview/afaa70dd57076e8ae4bab3fa3631ac20/1?pq-origsite=gscholar&cbl=18750&diss=y [accessed 2024-12-16]
37. Petkari E, Calvo A. Escaping psychodiagnostics: implementation of a Digital Educational Escape Room with future mental health professionals. Interact Learn Environ 2024 May 23:1-10. [doi: 10.1080/10494820.2024.2354429]
38. Bierbooms J, Feijt M, IJsselsteijn W, Sluis-Thiescheffer W, Weijmans M, van der Burg E, et al. Serious games for professional skills: the design of an escape room to explore the possibilities of eMental health. In: Proceedings of the Companion Publication of the 2020 ACM Designing Interactive Systems Conference. 2020 Presented at: DIS' 20 Companion; July 6-10, 2020; Eindhoven, The Netherlands. [doi: 10.1145/3393914.3395888]
39. Bierbooms J, Feijt MA, IJsselsteijn WA, Bongers IM. Design of a game-based training environment to enhance mental health care professionals' skills in using e-mental health: multiple methods user requirements analysis. JMIR Serious Games 2022 Jul 27;10(3):e34700 [FREE Full text] [doi: 10.2196/34700] [Medline: 35896032]
40. Pan R, Lo H, Neustaedter C. Collaboration, awareness, and communication in real-life escape rooms. In: Proceedings of the 2017 Conference on Designing Interactive Systems. 2017 Presented at: DIS '17; June 10-14, 2017; Edinburgh, UK. [doi: 10.1145/3064663.3064767]
41. Djaouti D, Alvarez J, Jessel JP. Classifying serious games: the G/P/S model. In: Handbook of Research on Improving Learning and Motivation through Educational Games: Multidisciplinary Approaches. Hershey, PA: IGI Global; 2011.
42. Stappers PJ, Giaccardi E. Research through design. In: Soegaard M, Friis-Dam R, editors. The Encyclopedia of Human-Computer Interaction. Aarhus, Denmark: The Interaction Design Foundation; 2014.
43. Zimmerman J, Forlizzi J, Evenson S. Research through design as a method for interaction design research in HCI. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2007 Presented at: CHI '07; April 28-May 3, 2007; San Jose, CA. [doi: 10.1145/1240624.1240704]
44. Jaarverslag GGzE 2020: werken aan mentale vooruitgang. GGzE. URL: https://www.jaarverslagggze.nl/ [accessed 2021-10-22]
45. Donetto S, Pierri P, Tsianakas V, Robert G. Experience-based co-design and healthcare improvement: realizing participatory design in the public sector. Des J 2015 May 07;18(2):227-248. [doi: 10.2752/175630615X14212498964312]
46. Jessen S, Mirkovic J, Ruland CM. Creating gameful design in mHealth: a participatory co-design approach. JMIR Mhealth Uhealth 2018 Dec 14;6(12):e11579 [FREE Full text] [doi: 10.2196/11579] [Medline: 30552080]
47. Robert G. Participatory action research: using experience-based co-design to improve the quality of healthcare services. In: Ziebland S, Coulter A, Calabrese JD, Locock L, editors. Understanding and Using Health Experiences: Improving Patient Care. Oxford, UK: Oxford University Press; 2013.
48. Boyatzis RE. Transforming Qualitative Information: Thematic Analysis and Code Development. Ann Arbor, MI: The University of Michigan; 1998.
49. Ogrinc G, Davies L, Goodman D, Batalden P, Davidoff F, Stevens D. SQUIRE 2.0 (Standards for QUality Improvement Reporting Excellence): revised publication guidelines from a detailed consensus process. BMJ Qual Saf 2016 Dec;25(12):986-992 [FREE Full text] [doi: 10.1136/bmjqs-2015-004411] [Medline: 26369893]
50. Connolly SL, Miller CJ, Lindsay JA, Bauer MS. A systematic review of providers' attitudes toward telemental health via videoconferencing. Clin Psychol (New York) 2020 Jan 06;27(2):10.1111/cpsp.12311 [FREE Full text] [doi: 10.1111/cpsp.12311] [Medline: 35966216]

51. Davies F, Shepherd HL, Beatty L, Clark B, Butow P, Shaw J. Implementing web-based therapy in routine mental health care: systematic review of health professionals' perspectives. J Med Internet Res 2020 Jul 23;22(7):e17362 [FREE Full text] [doi: 10.2196/17362] [Medline: 32706713]
52. Stoll J, Müller JA, Trachsel M. Ethical issues in online psychotherapy: a narrative review. Front Psychiatry 2019 Feb 11;10:993 [FREE Full text] [doi: 10.3389/fpsyt.2019.00993] [Medline: 32116819]
53. Ijaz A, Khan MY, Ali SM, Qadir J, Kamel Boulos MN. Serious games for healthcare professional training: a systematic review. Eur J Biomed Inform 2019;15(1):14-30 [FREE Full text]
54. Rooney P. Creating serious games at third level: evaluating the implications of an in-house approach. In: Proceedings of the European Conference on Game Based Learning. 2012 Presented at: ECGBL 2012; October 4-5, 2012; Cork, Ireland.
55. 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]
56. Moizer J, Lean J, Dell'Aquila E, Walsh P, Keary A, O'Byrne D, et al. An approach to evaluating the user experience of serious games. Comput Educ 2019 Jul;136:141-151. [doi: 10.1016/j.compedu.2019.04.006]
57. Pistono AM, dos Santos AM, Baptista RJ, Mamede HS. Framework for adaptive serious games. Comput Appl Eng Educ 2024 Mar 07;32(4):e22731. [doi: 10.1002/cae.22731]
58. Karavidas L, Apostolidis H, Tsiatsos T. Usability evaluation of an adaptive serious game prototype based on affective feedback. Information 2022 Sep 08;13(9):425. [doi: 10.3390/info13090425]

Abbreviations

EER: educational escape room
eMH: e-Mental health
GGzE: Geestelijke Gezondheidszorg Eindhoven en de Kempen
SDT: self-determination theory
SMI: severe mental illness
VR: virtual reality

Edited by A Mavragani; submitted 21.03.24; peer-reviewed by A Frey, IK Singgih, A Ferworn, E Vallefuoco; comments to author 10.05.24; revised version received 22.08.24; accepted 21.10.24; published 07.01.25.

Please cite as:
Bierbooms JJPA, Sluis-Thiescheffer WRJW, Feijt MA, Bongers IMB
Co-Design of an Escape Room for e-Mental Health Training of Mental Health Care Professionals: Research Through Design Study
JMIR Form Res 2025;9:e58650
URL: https://formative.jmir.org/2025/1/e58650
doi:10.2196/58650
PMID:

Original Paper

International Expert Consensus on Relevant Health and Functioning Concepts to Assess in Users of Tobacco and Nicotine Products: Delphi Study

Vivienne Law[1*], BSc; Esther F Afolalu[2*], PhD; Linda Abetz-Webb[3], MA; Lee Andrew Wemyss[4], BSc; Andrew Turner[4], BSc; Christelle Chrea[2], PhD

[1]Zebra26 Ltd., Devon, United Kingdom

[2]PMI R&D, Philip Morris Product S.A., Neuchâtel, Switzerland

[3]Patient-Centered Outcomes Assessments Ltd., Cheshire, United Kingdom

[4]DaSH Global, Cheshire, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Esther F Afolalu, PhD
PMI R&D, Philip Morris Product S.A.
Quai Jeanrenaud 5
Neuchâtel, 2000
Switzerland
Phone: 41 58 242 2218
Email: esther.afolalu@pmi.com

Abstract

Background: A Delphi study was conducted to reach a consensus among international clinical and health care experts on the most important health and functioning self-reported concepts when evaluating a switch from smoking cigarettes to using smoke-free tobacco and/or nicotine products (sf-TNPs).

Objective: The aim of this research was to identify concepts considered important to measure when assessing the health and functioning status of users of tobacco and/or nicotine products.

Methods: Experts (n=105), including health care professionals, researchers, and policy makers, from 26 countries with professional experience and knowledge of sf-TNPs completed a 3-round, adapted Delphi panel. Online surveys combining quantitative (MaxDiff best-worst scaling and latent class analysis) and qualitative assessments were used to rank and achieve alignment on the importance of 69 health and functioning concepts. All experts participating in round I completed round II, and 101 (95%) completed round III.

Results: The round I analysis identified 36 (52%) out of 69 concepts that were refined for the round II assessment. The highest-ranked concepts reflected health-related impacts, while the lowest-ranked ranked concepts were related to aesthetics and social impacts. Round II ranking reinforced the importance of concepts relating to health impacts, and the analysis resulted in 20 concepts retained for round III assessment. In round III, the 4 highest-ranked concepts were cardiovascular symptoms, shortness of breath, chest pain, and worry about smoking-related diseases and impact on general health, and they made up 50% of the total score in the MaxDiff analysis. Experts reported likelihood of seeing measurable levels of change in the final 20 concepts with a switch to an sf-TNP. The majority of experts felt it was “likely” or “extremely likely” to observe changes in concepts such as gum problems (74/101, 73%), phlegm or mucus while coughing or not coughing (72/101, 71%), general perception of well-being (72/101, 71%), and throat irritation or sore throat (72/101, 71%). Latent class analysis revealed subgroups of experts with different perceptions of the relative importance of the concepts, which varied depending on professional specialty and geographic region. For example, 74% (14/19) of oncologists aligned with the subgroup prioritizing physical health symptoms, while 71% (12/17) of experts from Asia aligned with the subgroup considering both physical health and psychosocial aspects.

Conclusions: This study identified key concepts to be considered in the development of a new measurement instrument to assess the self-reported health and functioning status of individuals using sf-TNPs. The findings contribute to the scientific evidence base for understanding and evaluating both the individual and public health impacts of sf-TNPs.

(JMIR Form Res 2025;9:e58614) doi:10.2196/58614

KEYWORDS

Delphi study; expert consensus; outcome measures; health and functioning; tobacco and/or nicotine products

Introduction

Reducing exposure to harmful and potentially harmful constituents in cigarette smoke by cessation of smoking or switching to reduced risk tobacco and/or nicotine products (TNPs) is a major public health focus worldwide [1]. Although nicotine-replacement therapy (NRT) and targeted counseling services can help individuals achieve their quitting goals, there can be hurdles to becoming and/or staying smoking free [2].

Smoke-free tobacco and/or nicotine products (sf-TNPs) refer to TNPs that do not undergo combustion and therefore do not produce the harmful smoke generated by cigarettes, which has been established as the main risk factor of tobacco-related diseases [3]. These sf-TNPs can include heated tobacco products, e-cigarettes (vapes), other e-vapor products (e-pipes, e-cigars, and so on), nicotine pouches, and smokeless tobacco products (snuff, snus, and chewing tobacco) [3]. Regulatory bodies, like the US Food and Drug Administration (FDA), require robust data to evaluate the health impacts of sf-TNPs and guide regulatory decisions on product authorization. For instance, the FDA's modified risk tobacco product pathway mandates comprehensive scientific evidence to support claims that a tobacco product reduces harm or the risk of tobacco-related disease in individual tobacco users and benefits the health of the population as a whole [4]. Some sf-TNPs have been recognized as modified-risk tobacco products under this pathway [4,5] and may help reduce the negative health effects associated with commercially marketed combustible tobacco products such as cigarettes [6-8].

The adverse effects of smoking on health have been well documented, whereas stopping smoking can improve health and reduce the risk of disease [6,9,10]. Evidence also suggests that switching away from cigarettes to using sf-TNPs can help reduce cigarette consumption and may lead to cessation in some cases [8]. While studies have shown that sf-TNPs reduce exposure to many of the harmful and potentially harmful constituents found in cigarette smoke, they may still pose health risks. For instance, e-cigarettes, heated tobacco products, and smokeless tobacco have been associated with some cardiovascular, respiratory, and oral health risks, albeit to a lesser extent than combustible cigarettes [11-14]. The long-term health effects of sf-TNPs are not yet fully understood, and further research is necessary to comprehensively assess their safety profile. In addition, little is known about the self-reported health impact of switching from cigarettes to sf-TNPs [6]. In this context, measuring the self-reported experience of health and functioning (including health status, functional status, and other health-related quality of life constructs) is crucial to understanding the impact of tobacco harm reduction strategies [15].

Generic health status measures, such as the Short Form Survey–12 [16] and Short Form Survey–36 instruments [17], have shown that those who smoke tend to report lower health status compared with those who never smoked, although the impact of cessation on health status seems to be more complex. Some smoking-specific measures have been developed but not yet widely standardized; therefore, they may not be directly applicable to sf-TNPs [9,18,19]. Crucially, current measures often lack the necessary sensitivity to detect longitudinal health changes in healthy populations due to high ceiling effects at baseline [20], hindering the assessment of changes in health and functioning following a switch from smoking cigarettes to using sf-TNPs.

To help address these challenges, the development of a new self-reported measure (ABOUT – Health and Functioning) was undertaken to assess the health and functioning status of individuals using sf-TNPs. This new measure is part of the portfolio of the ABOUT Toolbox [21], consisting of self-report measures to assess perceptions and behavioral outcomes related to the use of sf-TNPs. The initial preparatory phase of the instrument development was based on several research activities (systematic literature review, reanalysis of qualitative data, and expert insights [22]) and resulted in the identification of 69 health and functioning concepts relevant to TNP use. The qualitative research phase that followed consisted of (1) concept elicitation interviews of users to understand their perceptions of health and functioning after switching to sf-TNPs [23,24] and (2) the Delphi panel study reported in this paper. The aim of the Delphi study was to identify the health and functioning concepts considered most important by clinical and health care experts (ie, health care professionals, researchers, policy makers, and those involved in smoking cessation or tobacco harm reduction) when assessing the health and functioning status of individuals who stop smoking cigarettes or switch to using sf-TNPs or NRTs. Delphi panel methodology is a well-established process for determining consensus among relevant groups of individuals to aid issue prioritization, ranking, development, and obtaining agreement on guidelines, concept-framework development, and development of outcome measures [25-27], including issues relevant to smoking-related behaviors and sf-TNP use [25,28,29]. Delphi panels have previously been used to forecast trends and changes over time in health-related matters [30-32].

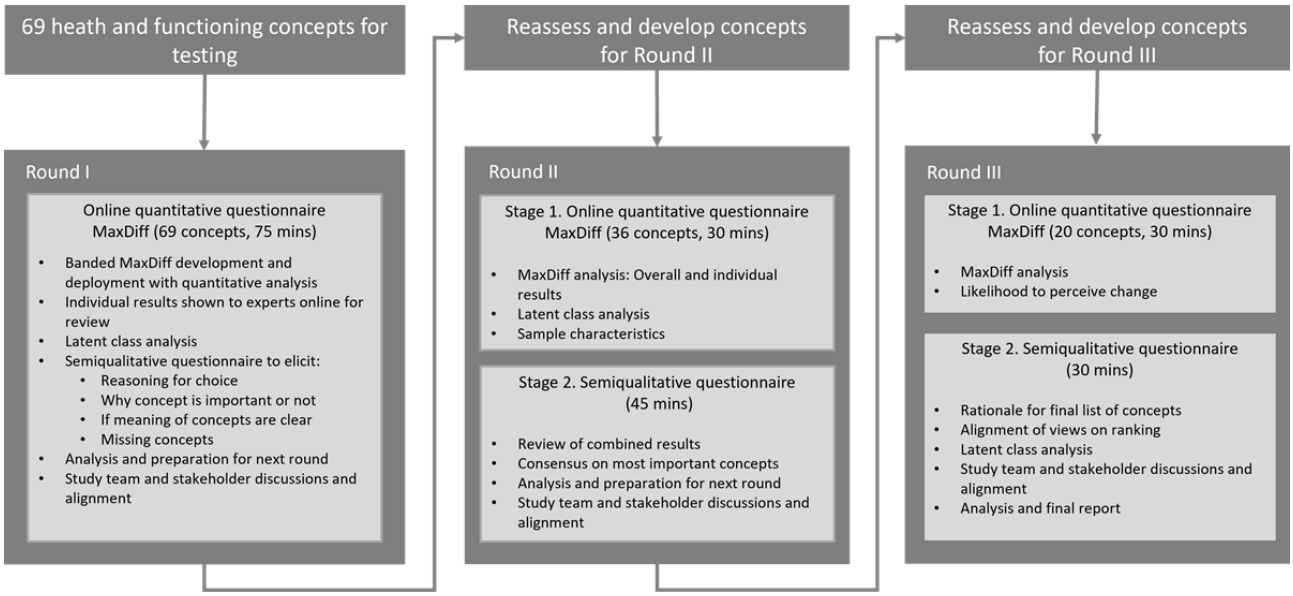
Methods

Objectives and Summary of Approach

This Delphi panel was organized for experts to select and rank health and functioning concepts considered important to assess when individuals stop smoking combustible TNPs (eg, cigarettes) or switch to using sf-TNPs (eg, e-cigarettes [vapes], heated tobacco, or smokeless tobacco products) or NRTs. The study was conducted and reported based on available guidance for conducting and reporting Delphi studies in health care research [33].

A preliminary set of 69 concepts was identified from earlier research activities including a scoping literature review, results from the reanalysis of previous TNP consumer qualitative research results, and expert opinion from a small group of key opinion leaders [22]. This list was presented for evaluation in this Delphi panel according to the process described in Figure 1. An adapted approach was used for the Delphi panel, whereby the results of round I were reviewed and refined in light of the ongoing qualitative concept elicitation interview findings from consumers of TNPs and ongoing review by key opinion leaders (Supplementary Information and Table S1 in Multimedia Appendix 1). The concept list to be evaluated in round II was adjusted accordingly and refined considering the round I findings. Round III was a detailed review of the final set of health and functioning concepts. This step also included an evaluation of likelihood of change that experts considered important for each concept for inclusion in a self-reported measure of clinical relevance.

Figure 1. Overview of the design and structure of the adapted Delphi panel study.



Participants and Recruitment

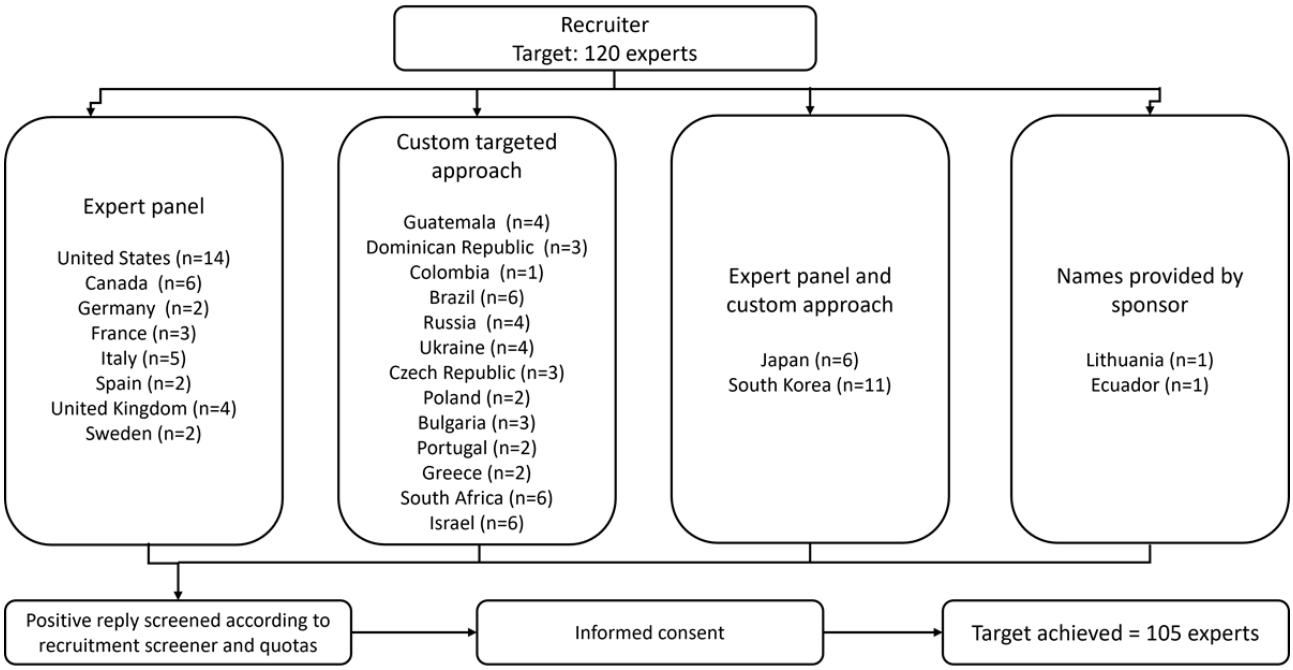
For this adapted Delphi panel, we recruited experts who routinely treat, communicate with, or advise individuals who smoke or wish to stop smoking or switch to sf-TNPs. The recruitment aim was to ensure that a minimum of 100 experts globally completed all 3 rounds of the Delphi panel. To anticipate attrition between rounds, the initial recruitment target was set at 120 experts.

Recruitment was undertaken by QualWorld, who identified experts through their survey panels, existing contacts, publications, word of mouth, and canvassing through appropriate organizations in individual countries. The process and route of contact to identify potential respondents are outlined in Figure 2.

To anticipate and account for potential sensitivities among health care practitioners about undertaking a project sponsored by a tobacco company, the recruitment company sent out an initial survey asking which industries experts would be interested in working on health-related topics. Tobacco companies were listed among other industries such as mining, pharmaceuticals, and oil and gas. Only experts stating unprompted that they would consider participating in studies by tobacco companies were chosen to receive further correspondence linked to the Delphi panel. Subsequent recruitment steps proceeded on an individual basis according to predefined inclusion and exclusion criteria (Textbox 1).

To ensure consistency in responses across different regions, the Delphi panel was conducted in English where possible and in local languages for experts in Japan, South Korea, Russia, Ukraine, Czech Republic, and Bulgaria. A professional translation service translated the survey materials and responses, and back-translation techniques were used to verify accuracy. The local bilingual recruitment team members also performed proofreading and quality checks from a language and contextual perspective to ensure the translated content accurately reflected the original meaning. For each step of the Delphi panel, the English and local language surveys were programmed and uploaded for a user acceptability test by the project team before launch.

Figure 2. Identification and recruitment of experts through specific channels for the adapted, 3-round Delphi panel.



Textbox 1. Inclusion and exclusion criteria for the participants recruited for the Delphi panel.

Inclusion criteria

- Had current clinical, academic, or professional experience in the area of smoking-related diseases, smoking cessation, nicotine addiction, or other smoking-related conditions, as well as health policy or advocacy work related to tobacco control or tobacco harm reduction
- Currently working in research, prescribing, or a recommendation capacity
- Knowledge of smoke-free tobacco and/or nicotine products or nicotine replacement therapies, works with patients to some degree, and/or recommends smoke-free tobacco and/or nicotine products or nicotine replacement therapies
- Expertise in one of the following groups:
  - Specialist physicians: smoking-related oncology; smoking-related respiratory disease, and smoking-related cardiovascular disease (specifically treatment of patients)
  - General practitioners and internal medicine physicians: those who see patients with a smoking-related condition or disease and recommend smoking cessation advice
  - Dentists or oral hygienists: those who see patients with smoking-related dental issues
  - Smoking cessation, addiction, or dependence: those who help users quit smoking with less harmful products or nicotine replacement therapy or worked with nicotine addiction; nurses undertaking smoking cessation activity; researchers in nicotine addictions; psychologists, social workers, or counselors
  - Health policy, advocacy, and non-profit organizations: smoking-related health policy, including tobacco harm reduction; organizations or charities that encourage people to quit; health policy and advocacy related to tobacco control and tobacco harm reduction. This can include physicians who work alongside the government and local policy as well as nonprofit organizations
- Ability and willingness to participate across the whole study
- Ability to complete the research in the English language (except for these countries: South Korea, Japan, Ukraine, Czech Republic, Russia, and Bulgaria)
- At least 30% of experts to identify as female (although this was not a strict quota for recruitment)

Exclusion criteria

- More than 50% of experts should not have been a consultant with or worked for the tobacco industry
- Had less than 5 years of clinical, academic, or professional experience
- Did not fall into one of the categories for inclusion
- Not willing to complete all 3 rounds of research

Ethical Considerations

The New England Independent Review Board reviewed the study protocol and all study materials and any amendments and granted approval of the study (reference 1 - 9184 - 1).

Before participation, all participants were required to review and sign an informed consent form, which provided a detailed description of the study, outlined the procedures, and specified participant expectations. Participants had the opportunity to ask questions and receive answers before signing the consent form.

All study data were deidentified and participants were identified only by unique ID numbers throughout the research rounds.

All experts were compensated and received a cash honorarium according to local professional rates in their respective countries, ranging from £115 to £450 (approximately US $147 to US $576 at the time of the research), for their time after each round.

Procedure and Analytical Methods

MaxDiff Analysis

This study incorporated the use of MaxDiff—a type of best-worst (maximum difference) scaling analytical method [34-36]. This methodology was primarily chosen due to the large number of initial concepts to be evaluated. Ranking and rating exercises with numerous concepts can lead to a cognitive burden, so a best-worst scaling approach overcomes fatigue while generating additional information by making respondents choose the most and least important concepts. The result is a list of concepts in order of relative importance [36-38].

For each round, the MaxDiff Sawtooth software algorithm [39] generated a list of scaled scores from most to least important, as well as the relative importance of one concept compared with another. This was based on the number of times that concept was rated as the most or least important. In addition, cumulative scores were used to differentiate the number of concepts accounting for 20%, 50%, and 70% of the most important concepts; the 50th percentile was used as an initial guide for which concepts to retain.

Rounds I and II also featured an “anchor” value, which was a score marking the boundary between concepts considered important (to be retained in the next round) or unimportant (to be excluded from the next round). This anchor value was derived from an anchor question representing a threshold of “important or not important” included for each concept to be assessed as above (ie, important) or below (ie, unimportant) the anchor. The anchor question in this survey required one of the following 3 responses: “none of these [items] are important,” “some of these [items] are important,” or “all of these [items] are important.” The anchor value was statistically derived using the Sawtooth software [38]. An illustration of a typical MaxDiff question is shown in Figure 3.

Figure 3. Typical presentation of the Delphi panel MaxDiff ranking question, including the anchor evaluation. NRT: nicotine replacement therapy; TNP: tobacco and/or nicotine products.

Which of the following statements/concepts are least and most important to measure and assess the impact on health and functioning when a TNP consumer switches from smoking a combustible TNP to using a smoke-free TNP or NRT?

Please choose one most important concept and one least important concept by clicking on the relevant button.

EXAMPLE [SINGLE ANSWER]

Most important	Choose one most important concept and one least important concept	Least important
◉	Shortness of breath or dyspnea	○
○	Cardiovascular symptoms (e.g. palpitations, blood pressure)	◉
○	Mouth ulcers	○
○	Physical endurance during everyday activities and exercise	○
○	Weight control	○

Considering only the concepts above, please indicate their overall importance:

Please click one answer

◉	None of these are important
○	Some of these are important
○	All of these are important

Latent Class Analysis

Latent class analysis [40] was also performed to determine whether some concepts were more important to some groups of experts over others, according to experts' demographic or other characteristics. This ensured that the final set of items was not biased toward one subset of health care professionals at the expense of others' preferences. Where there were concepts that were only important to subsets of the study sample but unimportant overall, these concepts could be retained.

Semiqualitative Assessment

Semiqualitative questions were incorporated into each Delphi panel round to provide an increased understanding of participants' responses to the results of the ranking and scoring exercise. These questions were asked to ensure no concepts were missing or needed rewording, and to understand the reasons behind experts' selections of importance or unimportance. They also sought to gauge agreement or disagreement with the overall ranking results and to gain insights into the reasons for these viewpoints. After each round, the semiqualitative comments were reviewed alongside the MaxDiff scores. This process involved coding and analyzing the qualitative feedback to identify common themes and insights that could provide context to the quantitative rankings. This feedback was then integrated with the quantitative MaxDiff results to refine and inform decisions about the list of concepts to be included for the next rounds. It helped explain the rationale behind the experts' choices and highlighted any discrepancies or areas of consensus, ensuring that both qualitative and quantitative perspectives were considered in the decision-making process.

Results

Descriptive Participant Demographics

Approximately 3100 experts were contacted, with a final sample of 105 experts recruited across 26 countries. Table 1 shows the demographics and participant characteristics, and Table 2 shows the participant quotas per region and per specialty. Of the initial 105 experts, all participated in exploratory round I and round II, and 101 (96%) completed round III (Table 2). Nearly all participants (90/105, 87%) reported frequently addressing smoking cessation with clients or patients. A total of 64% (67/105) of experts in the study identified as male, and most were located in the Americas, Europe, or Asia. Regarding place of work, three-quarters (79/105, 75%) of experts were based in hospitals or other clinical settings, mostly as medical specialists or general care practitioners.

Table 1. Demographics of Delphi panel participants (rounds I-III).

Demographic variable	Respondents	
	Rounds I and II (n=105), n (%)	Round III (n=101), n (%)
Sex		
Male	67 (64)	64 (63)
Female	38 (36)	37 (37)
Geographical region		
North America (United States and Canada)	20 (19)	17 (17)
Asia (Japan and South Korea)	17 (16)	17 (17)
Central, Caribbean, South America (Brazil, Colombia, Dominican Republic, and Guatemala)	15 (14)	15 (15)
Southern Europe (Spain, Portugal, Italy, Greece, and Bulgaria)	14 (13)	13 (13)
Western Europe (France, Germany, Sweden, United Kingdom, and the Netherlands)	13 (12)	13 (13)
Eastern Europe (Russia, Ukraine, and Lithuania)	9 (9)	9 (9)
Africa (South Africa)	6 (6)	5 (5)
Middle East (Israel)	6 (6)	6 (6)
Central Europe (Poland and Czech Republic)	5 (5)	5 (5)
Work setting		
Hospital	43 (41)	41 (41)
Clinical (other)	36 (34)	36 (36)
Professional	17 (16)	15 (15)
University hospital	5 (5)	5 (5)
University hospital; hospital setting	4 (4)	4 (4)
Main current role		
General practitioner or internal medicine	27 (26)	27 (27)
Oncology	19 (18)	19 (19)
Dentist or oral hygienist	16 (15)	15 (15)
Research	11 (10)	9 (9)
Cardiology	12 (11)	12 (12)
Respiratory	8 (8)	8 (8)
Counselor or psychologist	7 (7)	6 (6)
Advocacy or health policy or NGO[a]	5 (5)	5 (5)
Frequency of discussing smoking cessation with clients		
Frequent (at least weekly)	90 (86)	86 (85)
Not frequent (less than once a week)	11 (10)	11 (11)
Not applicable	4 (4)	4 (4)
Discussed the benefits of stopping smoking or provided information about sf-TNPs[b] or NRTs[c] to those who smoke and do not want or have difficulty quitting smoking		
Yes	103 (98)	99 (98)
No	0	0
Not available or prefer not to say	2 (2)	2 (2)[d,e]
Provided information on heat-not-burn or heated tobacco products (eg,*IQOS*, Glo) or tobacco vapor products (eg, Ploom Tech)		
Yes	50 (48)	46 (46)

Demographic variable	Respondents	
	Rounds I and II (n=105), n (%)	Round III (n=101), n (%)
No	53 (50)	53 (53)
Not available or prefer not to say	2 (2)	2 (2)
Provided information on smokeless tobacco (eg, Copenhagen Snuff, Swedish Match General Snus, Camel Snus, or any other local brands)		
Yes	39 (37)	35 (35)
No	64 (61)	64 (63)
Not available or prefer not to say	2 (2)	2 (2)
Provided information on e-cigarettes (eg, JUUL, Blu, Logic, or any other local brands)		
Yes	65 (62)	62 (61)
No	38 (36)	37 (37)
Not available or prefer not to say	2 (2)	2 (2)
Provided information on Nicotine replacement therapies (eg, nicotine gums, inhalers, nasal sprays, lozenges, patches)		
Yes	91 (87)	88 (87)
No	11 (10)	10 (10)
Not available or prefer not to say	3 (3)	3 (3)
Current or previous personal use of combustible TNPs[f]		
Never	65 (62)	62 (61)
Yes	35 (33)	34 (34)
Not available or prefer not to say	5 (5)	5 (5)
Current or previous personal use of sf-TNP or NRT of those that were currently or had used combustible TNPs		
Yes	12 (11)	12 (12)

[a]NGO: nongovernmental organization.

[b]sf-TNP: smoke-free tobacco or nicotine product.

[c]NRT: nicotine replacement therapy.

[d]Not available.

[e]Data missing for one respondent in round III, stage 2.

[f]TNP: tobacco or nicotine product.

Table 2. Participant quotas per region and per specialty for the start of round I of the Delphi panel (N=105).

Specialty	Asia	Africa	Middle East	North America	Central America	Southern Europe	Eastern Europe	Western Europe	Central Europe
Specialist physician	6	2	3	6	7	3	4	2	6
GP[a] or IM[b]	4	2	3	4	3	3	2	2	4
Dental specialist	3	1	0	2	2	2	3	1	2
SC[c] or DEP[d]	2	1	0	5	2	1	3	0	3
Health policy or NGO[e]	2	0	0	3	1	0	1	0	0
Total	17	6	6	20	14	9	13	5	15

[a]GP: general practitioner.

[b]IM: internal medicine specialist.

[c]SC: smoking cessation specialist.

[d]DEP: dependence or addiction specialist.

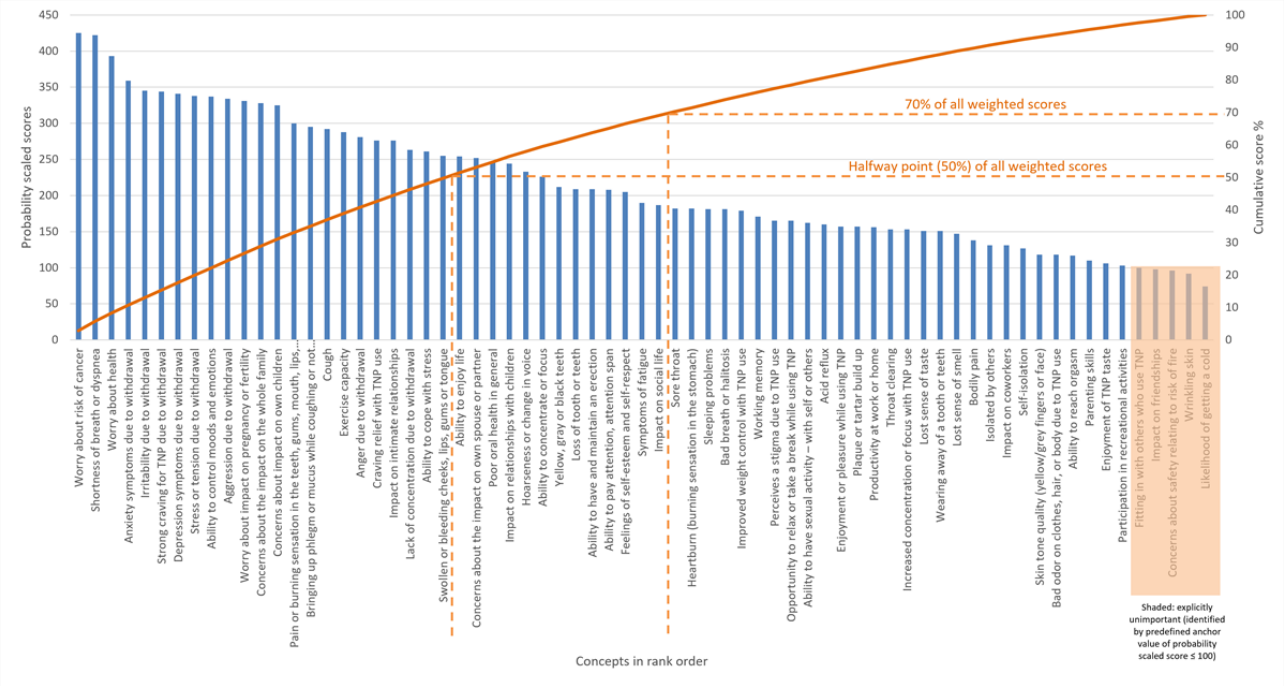
[e]NGO: nongovernmental organization.

MaxDiff Analysis and Semiqualitative Assessment for Reduction or Refinement of Health and Functioning Concepts

Round I

A cumulative score was calculated by adding all the scaled scores (in this case, the total score was 14,900.17) and dividing by the scaled score (concept 1: 425/14,900=3%) and adding to the scaled score of the next highest concept (concept 2: 422/14,900=3%+3% concept 1 and so on; Figure 4). Regarding the relative importance of concepts in round I, when switching from combustible to sf-TNPs, experts placed importance on withdrawal symptoms, general health impacts, and emotion-related impacts. Conversely, concepts that ranked low in priority were those that referred to aesthetic concerns that would have less of a health impact or were not considered exclusive to those who smoke. Experts considered withdrawal symptoms such as anxiety and irritability as potential barriers to switching to sf-TNPs and highlighted there were levels of satisfaction (eg, enjoyment and craving relief) that those who smoke cigarettes would not want to lose when switching to sf-TNPs

Figure 4. Scaling and ranking results for items considered in round I of the Delphi panel. TNP: tobacco and/or nicotine products.

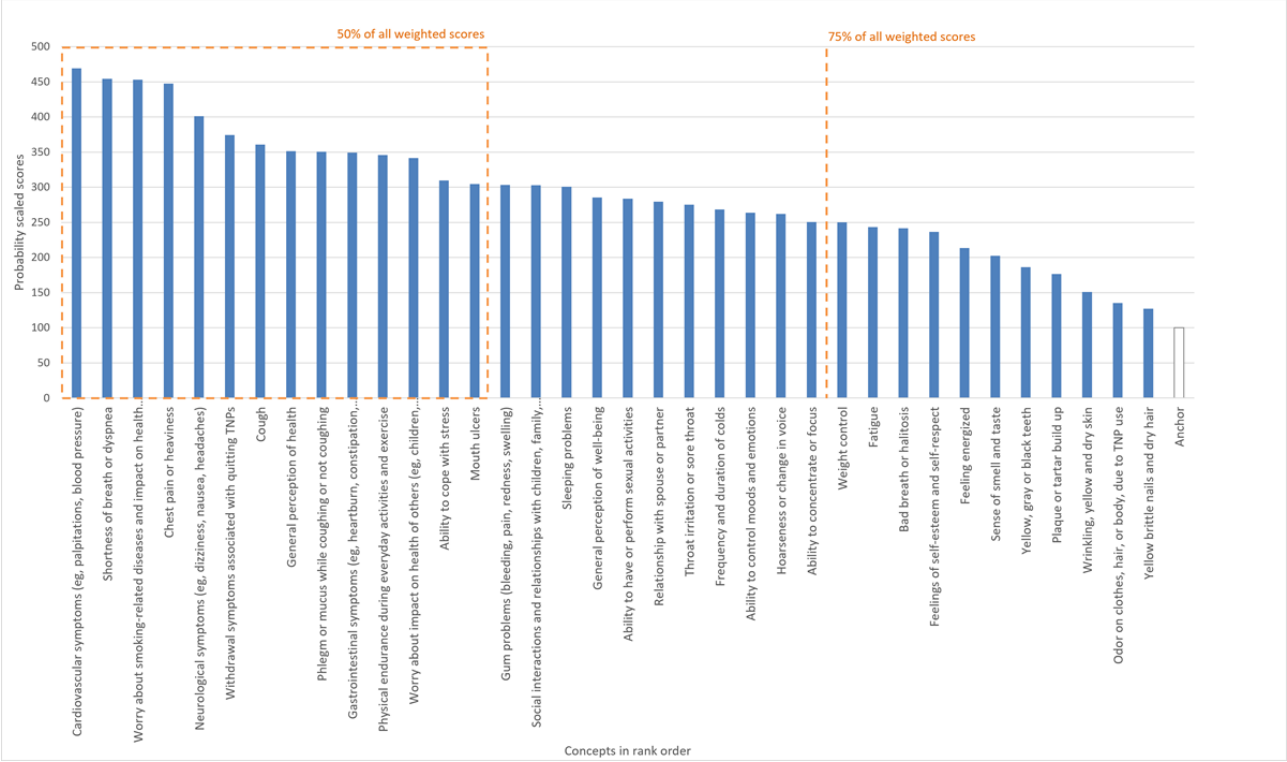


When determining which concepts to include in round II of the Delphi panel evaluation, the relative score (ie, importance) for each concept was considered, resulting in 36 out of 69 concepts (52% of all concepts in ranked order) accounting for 69% (10,285/14,900) of the total importance score. Based on the results and alignment with other activities ongoing within development of the health and functioning measurement instrument, the health and functioning concepts were further reviewed for relevance and clarity, and a different, refined, reduced, and synthesized list of 36 concepts was taken forward for evaluation in round II (refer to Supplementary Information and Table S1 in Multimedia Appendix 1 for further details).

Round II

Of the 36 concepts in this round, 14 made up 50% of the total scaled score (Figure 5). In general, physical health symptoms, and worries about the impact on health concepts remained the highest-ranked concepts and were considered by experts as more important than aesthetic considerations. Experts commented that concepts rated as unimportant and ranked lower in the logit analysis tended to lack specificity to smoking (ie, could be due to other causes), did not directly affect health, or were not reported frequently by TNP users. Of the 36 concepts evaluated in round II, a final selection of 20 highest-ranked concepts were identified for validation and discussion in round III.

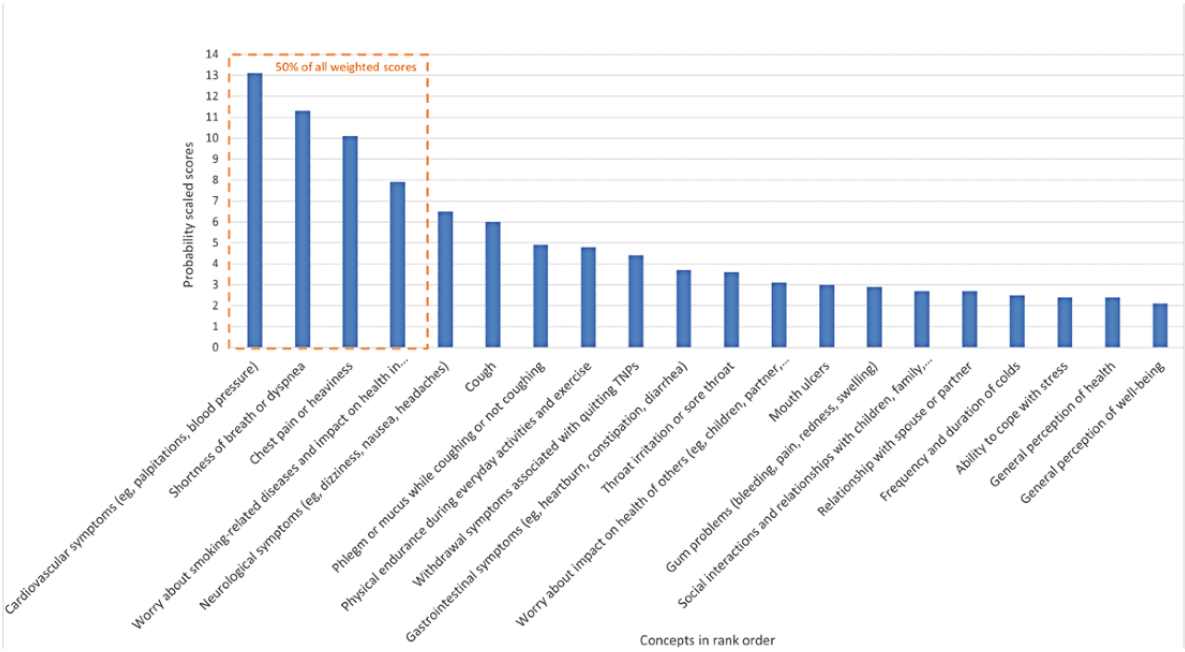
Figure 5. Scaling and ranking results for items considered in round II of the Delphi panel. TNP: tobacco and/or nicotine products.



Round III

Of the final 20 concepts ranked in round III (Figure 6), the sum of the top 4 concept scores accounted for more 50% of the total score: cardiovascular symptoms, shortness of breath or dyspnea, chest pain heaviness, and worry about smoking-related diseases and impact on health in general. Physical health concepts remained the most important experts across all 3 rounds

Figure 6. Ranking and ordering of the final 20 items identified in round III of the Delphi panel.



Latent Class Analysis

Latent class analysis of the results in round III revealed 2 identifiable subgroups, each with its own ranking and ordering of the final 20 concepts (Textbox 2). In both cases, concepts related to physical health, particularly cardiovascular and respiratory symptoms, ranked highest. In subgroup 1, the lowest-ranked concepts were related to psychosocial or general well-being and social interactions, and all scored equally low. For subgroup 2, there was minimal differentiation between concepts beyond the highest-ranking group.

Textbox 2. List of concepts in ranked order (highest to lowest) for the 2 subgroups identified by latent class analysis in round III of the Delphi panel.

Subgroup 1 in rank order

- Cardiovascular symptoms (eg, palpitations and blood pressure)
- Shortness of breath or dyspnea
- Chest pain or heaviness
- Neurological symptoms (eg, dizziness, nausea, and headaches)
- Cough
- Mouth ulcers
- Phlegm or mucus while coughing or not coughing
- Throat irritation or sore throat
- Gum problems (eg, bleeding, pain, redness, or swelling)
- Gastrointestinal symptoms (eg, heartburn, constipation, and diarrhea)
- Worry about smoking-related diseases and impact on health in general (eg, cancer and stroke)
- Withdrawal symptoms associated with quitting tobacco and/or nicotine products
- Frequency and duration of colds
- Physical endurance during everyday activities and exercise
- Ability to cope with stress
- Worry about impact on health of others (eg, children, partner, family, and friends)
- Relationship with spouse or partner
- General perception of health
- General perception of well-being
- Social interactions and relationships with children, family, friends, and colleagues

Subgroup 2 in rank order

- Cardiovascular symptoms (eg, palpitations and blood pressure)
- Shortness of breath or dyspnea
- Worry about smoking-related diseases and impact on health in general (eg, cancer and stroke)
- Chest pain or heaviness
- Physical endurance during everyday activities and exercise
- Neurological symptoms (eg, dizziness, nausea, and headaches)
- Cough
- Worry about impact on health of others (eg, children, partner, family, and friends)
- Social interactions and relationships with children, family, friends, and colleagues
- Withdrawal symptoms associated with quitting tobacco and/or nicotine products
- Relationship with spouse or partner
- Phlegm or mucus while coughing or not coughing
- General perception of health
- General perception of well-being
- Ability to cope with stress
- Gastrointestinal symptoms (eg, heartburn, constipation, and diarrhea)
- Throat irritation or sore throat
- Frequency and duration of colds
- Gum problems (eg, bleeding, pain, redness, and swelling)
- Mouth ulcers

In stage 2 of round III, following the MaxDiff analysis, the ranked concept lists for each of the 2 latent class subgroups were shown to the experts, who were asked to identify which group best reflected their own thinking and provide a rationale for their choice. Just over half of the experts (58/100, 58%) reported that subgroup 1 best reflected their thinking compared with (42/100, 42%) for subgroup 2. When asked about their rationale for aligning with subgroup 1, the main reason was a focus on the most important physical health symptoms (33/58, 57%). The main rationale for choosing subgroup 2 was a greater emphasis on the importance of the consumer and their perspective (10/42, 24%), including psychological elements, the well-being of consumers, and impact of smoking on the health of others (14/42, 33%).

Regarding the influence of professional specialty in self-alignment with subgroups (Table 3), preference for subgroup 1 tended to be among oncologists (14/19, 74%), dentists (11/15, 73%), counselors (4/6, 67%), and general practitioners or internal medicine specialists (17/27, 63%). Conversely, all policy advisors and charity workers (5/5) felt that subgroup 2 best reflected their views. When geographical location was considered (Table 3), there was a strong preference for subgroup 2 in participants from Asia (12/17, 71%) and a slight preference among those in North America (9/17, 53%). Experts from all other regions tended mostly to identify with subgroup 1 (Middle East: 5/6, 83%; Africa: 4/5, 80%; Central Europe: 4/5, 80%; Eastern Europe: 7/9, 78%; Western Europe: 9/13, 69%; Southern Europe: 8/13, 62%; and Central or South America: 8/15, 53%). Regarding the final selection of 20 concepts, both subgroups considered all concepts to be important and relevant for inclusion in any self-reported outcome measure.

Table 3. Region and specialty after self-identifying with a latent class subgroup in round III of the Delphi panel.

Experts' region and specialty	Total (n=100), n (%)[a]	Latent class subgroup 1, n (%)[b]	Latent class subgroup 2, n (%)[b]
Region	100 (100)	58 (58)	42 (42)
North America	17 (17)	8 (47)	9 (53)
Western Europe	13 (13)	9 (69)	4 (31)
Southern Europe	13 (13)	8 (62)	5 (38)
Africa	5 (5)	4 (80)	1 (20)
Central/South America	15 (15)	8 (53)	7 (47)
Central Europe	5 (5)	4 (80)	1 (20)
Asia	17 (17)	5 (29)	12 (71)
Eastern Europe	9 (9)	7 (78)	2 (22)
Middle East	6 (6)	5 (83)	1 (17)
Specialty	100 (100)	58 (58)	42 (42)
General practice or internal medicine	27 (27)	17 (63)	10 (37)
Dentist or oral hygienist	15 (15)	11 (73)	4 (27)
Oncologist	19 (19)	14 (74)	5 (26)
Cardiovascular specialist	11 (11)	5 (45)	6 (55)
Respiratory	8 (8)	3 (38)	5 (63)
Researcher	9 (9)	4 (44)	5 (56)
Counselor or psychologist	6 (6)	4 (67)	2 (33)
Policy advisor	4 (4)	0 (0)	4 (100)
Charity or advocacy	1 (1)	0 (0)	1 (100)

[a]Data for 100 experts were included, with one respondent (from South Africa) reporting not understanding the questions asked.

[b]Percentage use the n value in the "Total" column as the denominator.

Additional Semiqualitative Assessment—Likelihood to Perceive Change

In round III (in response to the question: "Now that you have seen the list of concepts collated from Rounds I & II, how likely do you think it is to see a change in this concept when a TNP user switches from a combustible TNP to an sf-TNP?"), most experts reported they would expect to find measurable levels of change in 13 out of 20 individual concepts. Furthermore, over half of the respondents stated it was "likely" or "extremely likely" that a measurable change would occur in these concepts when a TNP user switched from a combustible TNP to a sf-TNP (Figure 7). For example, majority of experts felt that it was "likely" or "extremely likely" to observe a change in gum problems (74/101, 73%), phlegm or mucus while coughing or not coughing (72/101, 71%), general perception of well-being (72/101, 71%), and throat irritation/sore throat (72/101, 71%). In addition, between 5% and 20% considered that switching

was “unlikely” or “very unlikely” to result in a change in the severity of any single concept, including cardiovascular symptoms, withdrawal symptoms associated with quitting, shortness of breath, and worry about smoking-related diseases. Reasons given by experts for considering a change in a concept to be unlikely are listed in Table 4 and were mostly related to lack of perceived substantial differences between combustible TNPs and sf-TNPs.

Figure 7. Experts’ opinions (percentage of experts) regarding likelihood of observing change in the severity of the final 20 items in round III of the Delphi panel.

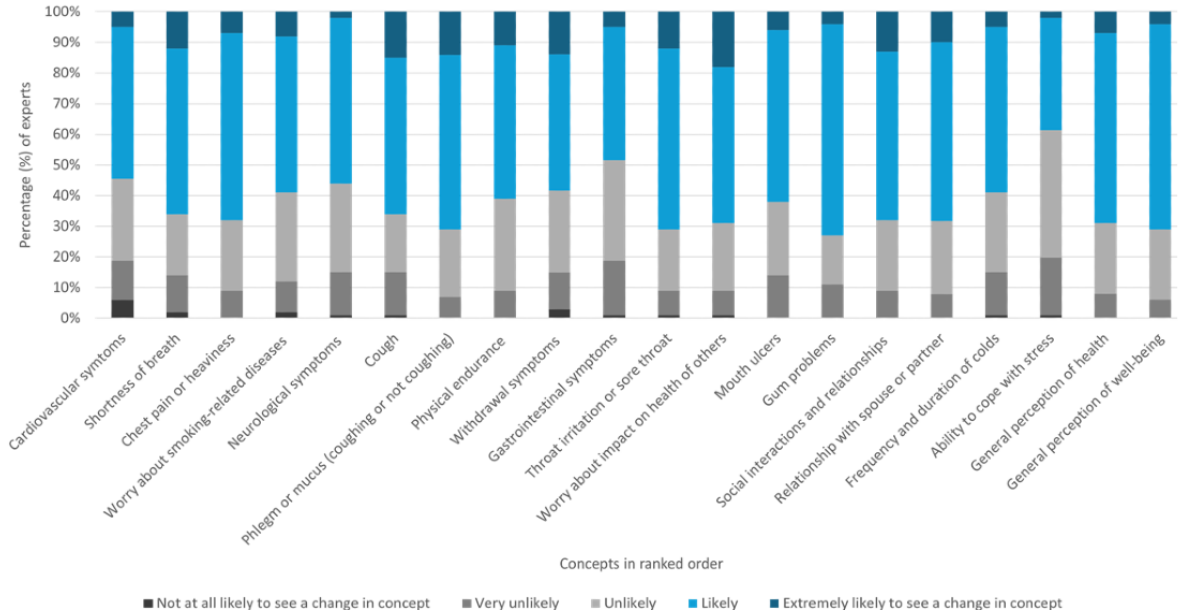


Table 4. Reasons given by experts for considering a change in a concept to be unlikely for the final 20 items in round III of the Delphi panel.

Concept (number of experts responding it was unlikely to see a change)	Rationale reported by experts
Cardiovascular symptoms (n=6)	• Unlikely to see substantial modifications between the combustible TNPs[a] and sf-TNPs[b] as tobacco or nicotine can cause cardiovascular disease • Active ingredients are the same and present in both combustible TNPs and sf-TNPs products
Shortness of breath or dyspnea (n=2)	• No change or no substantial modifications in switching to sf-TNPs
Worry about smoking-related diseases and impact on health in general (eg, cancer and stroke) (n=2)	• Symptoms will occur with using both combustible TNPs and sf-TNPs
Worry about impact on health of others (eg, children, partner, family, and friends) (n=1)	• Other dangerous chemicals in sf-TNPs
Frequency and duration of colds (n=1)	• Colds are unlikely to change as they are a nuisance and not a major health threat

[a]TNPs: tobacco and/or nicotine products.

[b]sf-TNPs: smoke-free tobacco and/or nicotine products.

Discussion

Principal Findings

This adapted Delphi panel study was conducted to identify the health and functioning concepts that experts consider to be most relevant and useful for incorporation into a new TNP-specific health and functioning measurement instrument to assess the self-reported impact of switching from combustible TNPs to sf-TNPs [21]. We recruited a geographically and professionally representative panel of experts who routinely treated those who smoke or communicated on smoking and TNP use. They were asked to reduce and refine a preliminary list of health and functioning concepts and to rank the final list in order of importance. After 3 rounds of investigation, the initial list of 69 concepts was refined to the most important 20 they considered could be included in a new self-reported measure (Table 5). The range of final concepts reflected health and psychosocial issues associated with smoking, including respiratory and cardiovascular health, withdrawal symptoms, and worries about the impact on health of self and others.

Table 5. Concepts recommended and agreed after each round of the Delphi panel.

Rank	Round I (n=69), ranked in order of importance (highest to lowest)	Round II (n=36), ranked in order of importance (highest to lowest)	Round III (n=20), ranked in order of importance (highest to lowest)
1	Worry about risk of cancer	Cardiovascular symptoms (eg, palpitations, blood pressure)	Cardiovascular symptoms (eg, palpitations, blood pressure)
2	Shortness of breath or dyspnea	Shortness of breath or dyspnea	Shortness of breath or dyspnea
3	Worry about health	Worry about smoking-related diseases and impact on health in general (eg, cancer, stroke)	Chest pain or heaviness
4	Anxiety symptoms due to withdrawal	Chest pain or heaviness	Worry about smoking-related diseases and impact on health in general (eg, cancer, stroke)
5	Irritability due to withdrawal	Neurological symptoms (eg, dizziness, nausea, headaches)	Neurological symptoms (eg, dizziness, nausea, headaches)
6	Strong craving for TNP[a] due to withdrawal	Withdrawal symptoms associated with quitting TNPs	Cough
7	Depression symptoms due to withdrawal	Cough	Phlegm or mucus while coughing or not coughing
8	Stress or tension due to withdrawal	General perception of health	Physical endurance during everyday activities and exercise
9	Ability to control moods and emotions	Phlegm or mucus while coughing or not coughing	Withdrawal symptoms associated with quitting TNPs
10	Aggression due to withdrawal	Gastrointestinal symptoms (eg, heartburn, constipation, diarrhea)	Gastrointestinal symptoms (eg, heartburn, constipation, diarrhea)
11	Worry about impact on pregnancy or fertility	Physical endurance during everyday activities and exercise	Throat irritation or sore throat
12	Concerns about the impact on the whole family	Worry about impact on health of others (eg, children, partner, family, friends)	Worry about impact on health of others (eg, children, partner, family, friends)
13	Concerns about impact on own children	Ability to cope with stress	Mouth ulcers
14	Pain or burning sensation in the teeth, gums, mouth, lips, throat, or tongue	Mouth ulcers	Gum problems (bleeding, pain, redness, swelling)
15	Bringing up phlegm or mucus while coughing or not coughing	Gum problems (bleeding, pain, redness, swelling)	Social interactions and relationships with children, family, friends, colleagues
16	Cough	Social interactions and relationships with children, family, friends, colleagues	Relationship with spouse or partner
17	Exercise capacity	Sleeping problems	Frequency and duration of colds
18	Anger due to withdrawal	General perception of well-being	Ability to cope with stress
19	Craving relief with TNP use	Ability to have or perform sexual activities	General perception of health
20	Impact on intimate relationships	Relationship with spouse or partner	General perception of well-being
21	Lack of concentration due to withdrawal	Throat irritation or sore throat	—
22	Ability to cope with stress	Frequency and duration of colds	—
23	Swollen or bleeding cheeks, lips, gums, or tongue	Ability to control moods and emotions	—
24	Ability to enjoy life	Hoarseness or change in voice	—
25	Concerns about the impact on own spouse or partner	Ability to concentrate or focus	—
26	Poor oral health in general	Weight control	—
27	Impact on relationships with children	Fatigue	—
28	Hoarseness or change in voice	Bad breath or halitosis	—
29	Ability to concentrate or focus	Feelings of self-esteem and self-respect	—
30	Yellow, gray, or black teeth	Feeling energized	—
31	Loss of tooth or teeth	Sense of smell and taste	—

Rank	Round I (n=69), ranked in order of importance (highest to lowest)	Round II (n=36), ranked in order of importance (highest to lowest)	Round III (n=20), ranked in order of importance (highest to lowest)
32	Ability to have and maintain an erection	Yellow, gray, or black teeth	—
33	Ability to pay attention, attention span	Plaque or tartar build up	—
34	Feelings of self-esteem and self-respect	Wrinkling, yellow, and dry skin	—
35	Symptoms of fatigue	Odor on clothes, hair, or body, due to TNP use	—
36	Impact on social life	Yellow brittle nails and dry hair	—
37	Sore throat	—	—
38	Heartburn (burning sensation in the stomach)	—	—
39	Sleeping problems	—	—
40	Bad breath or halitosis	—	—
41	Improved weight control with TNP use	—	—
42	Working memory	—	—
43	Perceives a stigma due to TNP use	—	—
44	Opportunity to relax or take a break while using TNP	—	—
45	Ability to have sexual activity—with self or others	—	—
46	Acid reflux	—	—
47	Enjoyment or pleasure while using TNP	—	—
48	Plaque or tartar build up	—	—
49	Productivity at work or home	—	—
50	Throat clearing	—	—
51	Increased concentration or focus with TNP use	—	—
52	Lost sense of taste	—	—
53	Wearing away of a tooth or teeth	—	—
54	Lost sense of smell	—	—
55	Bodily pain	—	—
56	Isolated by others	—	—
57	Impact on co-workers	—	—
58	Self-isolation	—	—
59	Skin tone quality (yellow or gray fingers or face)	—	—
60	Bad odor on clothes, hair, or body due to TNP use	—	—
61	Ability to reach orgasm	—	—
62	Parenting skills	—	—
63	Enjoyment of TNP taste	—	—
64	Participation in recreational activities	—	—
65	Fitting in with others who use TNP	—	—
66	Impact on friendships	—	—
67	Concerns about safety relating to risk of fire	—	—
68	Wrinkling skin	—	—
69	Likelihood of getting a cold	—	—

[a]TNP: tobacco and/or nicotine products.

Overall, the final 20 concepts were considered both clinically relevant and important to the experts in their evaluation of the possible impact of switching to sf-TNPs or stopping smoking cigarettes. We observed that the concepts consistently ranked highly among experts focused on objectively measurable health consequences such as respiratory and cardiovascular symptoms,

as well as overall physical functioning. In contrast, concepts such as the ability to cope with stress, risk of getting colds, oral health, physical appearance, social functioning, and sensory impacts (taste, sense of smell, and odor) were ranked as less important. Those concepts ranking highest reflect current evidence regarding smoking-related health outcomes and potential improvements attainable by ceasing cigarette consumption or switching to sf-TNPs [6,41-46]. The experts' knowledge of evidence for the effect of cessation or switching on pre-existing respiratory conditions [6,43,44,47,48] may also be reflected in the ranking.

It is essential to note that the relative ranking of these concepts is likely to be different for TNP consumers compared with health care professionals. Individuals trying to stop smoking or switch to sf-TNPs recognize the respiratory and other physical health and functional benefits and understand the potential for improved quality of life by reducing cigarette consumption or stopping smoking [49,50], as well as reduced exposure of others to smoking-related harm [6,51]. They also tend to report benefits and preferences for general hygiene and smell, better oral health, and sensory improvements [6,41,43]. However, no concepts related to general physical appearance or hygiene were in the final top-ranking concepts of the current study. These results explicitly reflect the participants' own experiences and opinions, which may differ from those of TNP users and be dependent on TNP user characteristics.

Although objective health outcomes such as cardiovascular and respiratory function were rated highly by experts throughout the study, some individuals rated psychosocial outcomes and effects on families and others (through secondhand smoke and so on) as equally important as objective physical measures. Latent class analysis in the final round of the study suggested that geographical location, cultural milieu, and professional specialism may each play a major role in these observations. Specifically, we identified a subgroup that focused mostly on physical, objectively measurable concepts (eg, cardiovascular symptoms, shortness of breath, chest pain, neurological symptoms, and cough), compared with the other subgroup, which considered both objective and subjective or emotional concepts (eg, worry about smoking-related diseases and impact on health both of self and others) to be of equal importance. These results may correspond with established patterns of perceptions and considerations among Western cultures, which may be broadly defined as individualistic (ie, driven by personal goals), whereas those from Asia and South America can be defined as collectivist (reflecting the primacy of mutual obligations among members of society) [51-53]. Furthermore, qualitative investigation of alignment with the subgroups revealed that for certain job roles (eg, counselors and policy advisors), the consumer-focused concepts were more important than for other health care professionals (eg, specialist clinicians), who were primarily concerned with objective, physical concepts. These findings have important implications for the development of a self-report measure that can be widely used and is adaptable to individual or local requirements [54]. When disseminating results based on the newly developed self-reported measure, it may be essential to tailor communications on sf-TNPs for specific audiences with different considerations of what is clinically meaningful to the experts themselves, as well as to the individual TNP user [26,33,54]. To validate the latent class findings in round III, we asked respondents to self-identify with the most relevant of the 2 subgroups. Overall, participants were able to do so readily, indicating that the 2 classes were culturally and professionally relevant, and that the new self-report measure will maximize content validity by including the range of concepts identified and ranked in round III to accommodate cultural and professional requirements and differences.

There are 3 main strengths of this study. First, we used the MaxDiff ranking to address the challenge of ranking an initial high number of concepts. Second, the incorporation of an anchor was essential in objectively identifying concepts for elimination. Third, a large panel of experts representing a broad range of professional expertise and geographies was recruited for this study. This allowed us to confirm validity of the concepts identified in the preparatory phase of the development of the new measure. It also enabled us to parse out cultural and professional subtleties that are important to consider when disseminating outcomes from the new measure. Development of an outcome measure involving both the target population and clinicians is a key component of creating a high-quality instrument [33,54]. Building this manner of collaboration and bridging between TNP users and relevant experts in this Delphi panel into the development of this new instrument would serve to enhance the potential quality and validity of the final measure by feeding into subsequent qualitative and psychometric evaluations [26].

There are also some limitations to our study. First, we did not use a strict Delphi panel process, and round I may be seen as a preliminary evaluation rather than a pure ranking and scoring exercise. Delphi panel methodology typically follows a highly structured process. In this particular case, it would have been ideal to perform additional preliminary qualitative insight work to review all the elements of a study, formally review and refine concept terminology, and test with external audiences. And, as previously mentioned, this Delphi panel focused on the experience and perceptions of experts and professionals who interact with TNP users. Consequently, the findings reported here do not have the vital context of TNP consumers opinions and perceptions. Unfortunately, it would have been impractical to address both angles in a single study with an initial list of 69 concepts. Instead, further research should consider evaluating TNP users' perceptions and rankings in a separate study using similar methodology. In addition, there is a potential for selection bias in the study due to the exclusion of experts who specified they would not consider participating in studies by tobacco companies. This exclusion could result in a sample that is not fully representative of the broader expert community, potentially impacting the generalizability of the findings. To mitigate this, we used a diverse recruitment strategy to include experts from various regions and specialties. However, the views of nonparticipating experts might differ from those who participated, and this limitation should be considered when interpreting the results. Finally, the survey was conducted in various languages, and despite the measures taken to ensure the consistency and validity of the translated surveys, linguistic and cultural differences may have led to variations in interpretation

of the survey. These variations could potentially affect the study's findings and should be considered when interpreting the results.

Conclusion

In conclusion, this 3-round, adapted Delphi panel identified a ranked list of 20 concepts to be considered when assessing the health and functioning status of individuals who stop smoking cigarettes or switch to using sf-TNPs or NRTs. It is to be expected that the scale of importance of each concept will vary based on the health status and concerns of an individual, whereas the ranking presented here represents a global, generalized view provided by the participating experts. In addition, the sensitivity of the concepts to accurately reflect changes in TNP use behavior will need to be determined. This would support the evaluation of the self-reported experience and impact of switching from conventional cigarettes to sf-TNPs on health risks and contribute to the regulatory and scientific evidence base for understanding both the individual and public health impacts of sf-TNPs.

Acknowledgments

We wish to thank David Floyd for providing editorial assistance. Philip Morris International is the sole source of funding and sponsor of this research.

Data Availability

The datasets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

The authors declare the following financial interests or personal relationships which may be considered as potential conflicting interests: EA and CC are employees of Philip Morris Products S.A. VL, LAW, LW, and AT were contracted and paid by Philip Morris Products S.A.

Multimedia Appendix 1
Amended/modified Delphi panel approach.
[DOCX File , 38 KB - formative_v9i1e58614_app1.docx]

References

1. WHO report on the global tobacco epidemic 2021: addressing new and emerging products. Geneva, Switzerland: World Health Organization (WHO); 2021.
2. Mersha AG, Gould GS, Bovill M, Eftekhari P. Barriers and facilitators of adherence to nicotine replacement therapy: a systematic review and analysis using the capability, opportunity, motivation, and behaviour (COM-B) model. Int J Environ Res Public Health 2020;17(23):8895 [FREE Full text] [doi: 10.3390/ijerph17238895] [Medline: 33265956]
3. Food and Drug Administration (FDA). 2020. URL: https://www.fda.gov/tobacco-products/products-guidance-regulations/products-ingredients-components [accessed 2021-12-21]
4. Guidance for Industry - Modified Risk Tobacco Product Applications (Draft Guidance). Silver Spring, MD: Food and Drug Administration (FDA) Center for Tobacco Products; 2012.
5. Modified risk orders. Food and Drug Administration (FDA). URL: https://www.fda.gov/tobacco-products/advertising-and-promotion/modified-risk-orders [accessed 2021-12-21]
6. Glasser AM, Collins L, Pearson JL, Abudayyeh H, Niaura RS, Abrams DB, et al. Overview of electronic nicotine delivery systems: a systematic review. Am J Prev Med 2017;52(2):e33-e66. [doi: 10.1016/j.amepre.2016.10.036] [Medline: 27914771]
7. Hajek P, Phillips-Waller A, Przulj D, Pesola F, Smith KM, Bisal N, et al. E-cigarettes compared with nicotine replacement therapy within the UK stop smoking services: the TEC RCT. Health Technol Assess 2019;23(43):1-82 [FREE Full text] [doi: 10.3310/hta23430] [Medline: 31434605]
8. McRobbie HJ, Phillips-Waller A, El Zerbi C, McNeill A, Hajek P, Pesola F, et al. Nicotine replacement treatment, e-cigarettes and an online behavioural intervention to reduce relapse in recent ex-smokers: a multinational four-arm RCT. Health Technol Assess 2020;24(68):1-82 [FREE Full text] [doi: 10.3310/hta24680] [Medline: 33270009]
9. Goldenberg M, Danovitch I, IsHak WW. Quality of life and smoking. Am J Addict 2014;23(6):540-562. [doi: 10.1111/j.1521-0391.2014.12148.x] [Medline: 25255868]
10. 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. Lancet 2020;396(10248):413-446 [FREE Full text] [doi: 10.1016/S0140-6736(20)30367-6] [Medline: 32738937]
11. Qasim H, Karim ZA, Rivera JO, Khasawneh FT, Alshbool FZ. Impact of electronic cigarettes on the cardiovascular system. J Am Heart Assoc 2017;6(9):e006353 [FREE Full text] [doi: 10.1161/JAHA.117.006353] [Medline: 28855171]
12. Critchley JA, Unal B. Health effects associated with smokeless tobacco: a systematic review. Thorax 2003;58(5):435-443 [FREE Full text] [doi: 10.1136/thorax.58.5.435] [Medline: 12728167]

13. Znyk M, Jurewicz J, Kaleta D. Exposure to heated tobacco products and adverse health effects, a systematic review. Int J Environ Res Public Health 2021;18(12):6651 [FREE Full text] [doi: 10.3390/ijerph18126651] [Medline: 34205612]
14. Marques P, Piqueras L, Sanz M. An updated overview of e-cigarette impact on human health. Respir Res 2021;22(1):151 [FREE Full text] [doi: 10.1186/s12931-021-01737-5] [Medline: 34006276]
15. Institute of Medicine, Board on Population Health and Public Health Practice, Committee on Scientific Standards for Studies on Modified Risk Tobacco Products. Scientific Standards for Studies on Modified Risk Tobacco Products. Washington, DC: National Academy of Science; 2012.
16. Gasperini B, Barbadoro P, Cherubini A, Pierri F, D'Errico MM, Di Stanislao F, et al. The earlier the better: health-related quality of life is inversely related to age of smoking cessation in older former Italian smokers. Aging Clin Exp Res 2017;29(4):655-663. [doi: 10.1007/s40520-016-0608-4] [Medline: 27456679]
17. Sarna L, Bialous SA, Cooley ME, Jun H, Feskanich D. Impact of smoking and smoking cessation on health-related quality of life in women in the nurses' health study. Qual Life Res 2008;17(10):1217-1227. [doi: 10.1007/s11136-008-9404-8] [Medline: 18931942]
18. Kulasekaran A, Proctor C, Papadopoulou E, Shepperd CJ, Guyer R, Gandek B, et al. Preliminary evaluation of a new German translated tobacco quality of life impact tool to discriminate between healthy current and former smokers and to explore the effect of switching smokers to a reduced toxicant prototype cigarette. Nicotine Tob Res 2015;17(12):1456-1464. [doi: 10.1093/ntr/ntv024] [Medline: 25914263]
19. Olufade AO, Shaw JW, Foster SA, Leischow SJ, Hays RD, Coons SJ. Development of the smoking cessation quality of life questionnaire. Clin Ther 1999;21(12):2113-2130. [doi: 10.1016/s0149-2918(00)87242-2] [Medline: 10645757]
20. Ware JE, Gandek B, Kulasekaran A, Guyer R. Evaluation of smoking-specific and generic quality of life measures in current and former smokers in Germany and the United States. Health Qual Life Outcomes 2015;13:128 [FREE Full text] [doi: 10.1186/s12955-015-0316-3] [Medline: 26276447]
21. Chrea C, Acquadro C, Afolalu EF, Spies E, Salzberger T, Abetz-Webb L, et al. Developing fit-for-purpose self-report instruments for assessing consumer responses to tobacco and nicotine products: the ABOUT™ toolbox initiative. F1000Res 2018;7:1878 [FREE Full text] [doi: 10.12688/f1000research.16810.1] [Medline: 30906527]
22. Afolalu EF, Spies E, Bacso A, Clerc E, Abetz-Webb L, Gallot S, et al. Impact of tobacco and/or nicotine products on health and functioning: a scoping review and findings from the preparatory phase of the development of a new self-report measure. Harm Reduct J 2021;18(1):79 [FREE Full text] [doi: 10.1186/s12954-021-00526-z] [Medline: 34330294]
23. Afolalu EF, Spies E, Bacso A, Clerc E, Abetz-Webb L, Gallot S, et al. The process of developing a new outcome measure to assess the impact of tobacco and/or nicotine-containing products on health and functioning. 2020 Presented at: 26th Annual Meeting Society for Research on Nicotine and Tobacco (SRNT); March 11-14, 2020; New Orleans, LA.
24. Afolalu EF, Lanar S, Abetz-Webb L, Bacso A, Blankenspoor M, Koochaki P, et al. Qualitative health and functioning concept elicitation interviews in Japanese users of heated tobacco products. 2021 Presented at: 27th Annual Meeting Society for Research on Nicotine and Tobacco (SRNT); February 24, 2021; Virtual event URL: https://www.pmiscience.com/content/dam/pmiscience/en/publications/presentations-and-posters/pdf/srnt-2021-poster-esther-afolalu.pdf
25. Asfar T, Schmidt M, Ebrahimi Kalan M, Wu W, Ward KD, Nakkash RT, et al. Delphi study among international expert panel to develop waterpipe-specific health warning labels. Tob Control 2020;29(2):159-167 [FREE Full text] [doi: 10.1136/tobaccocontrol-2018-054718] [Medline: 30696784]
26. Hepworth LR, Rowe FJ. Using delphi methodology in the development of a new patient-reported outcome measure for stroke survivors with visual impairment. Brain Behav 2018;8(2):e00898 [FREE Full text] [doi: 10.1002/brb3.898] [Medline: 29484260]
27. Santaguida PL, Oliver D, Gilsing A, Lamarche L, Griffith LE, Mangin D, et al. Delphi consensus on core criteria set selecting among health-related outcome measures (HROM) in primary health care. J Clin Epidemiol 2020;127:105-116. [doi: 10.1016/j.jclinepi.2020.04.028] [Medline: 32417261]
28. Blaser J, Cornuz J. Experts' consensus on use of electronic cigarettes: a Delphi survey from Switzerland. BMJ Open 2015;5(4):e007197 [FREE Full text] [doi: 10.1136/bmjopen-2014-007197] [Medline: 25877274]
29. O'Neill N, Dogar O, Jawad M, Kellar I, Kanaan M, Siddiqi K. Which behavior change techniques may help waterpipe smokers to quit? An expert consensus using a modified Delphi technique. Nicotine Tob Res 2018;20(2):154-160 [FREE Full text] [doi: 10.1093/ntr/ntw297] [Medline: 27798091]
30. Lintonen T, Konu A, Rönkä S, Kotovirta E. Drugs foresight 2020: a Delphi expert panel study. Subst Abuse Treat Prev Policy 2014;9:18 [FREE Full text] [doi: 10.1186/1747-597X-9-18] [Medline: 24885142]
31. Marten O, Koerber F, Bloom D, Bullinger M, Buysse C, Christensen H, et al. A Delphi study on aspects of study design to overcome knowledge gaps on the burden of disease caused by serogroup B invasive meningococcal disease. Health Qual Life Outcomes 2019;17(1):87 [FREE Full text] [doi: 10.1186/s12955-019-1159-0] [Medline: 31118091]
32. O'Donnell S, Doyle G, O'Malley G, Browne S, O'Connor J, Mars M, et al. Establishing consensus on key public health indicators for the monitoring and evaluating childhood obesity interventions: a Delphi panel study. BMC Public Health 2020;20(1):1733 [FREE Full text] [doi: 10.1186/s12889-020-09814-y] [Medline: 33203390]

33. Jünger S, Payne SA, Brine J, Radbruch L, Brearley SG. Guidance on conducting and reporting DElphi studies (CREDES) in palliative care: recommendations based on a methodological systematic review. Palliat Med 2017;31(8):684-706. [doi: 10.1177/0269216317690685] [Medline: 28190381]
34. Strasser A. Design and evaluation of ranking-type Delphi studies using best-worst-scaling. Technol Anal Strateg Manag 2018;31(4):492-501. [doi: 10.1080/09537325.2018.1521956]
35. Uy EJB, Bautista DC, Xin X, Cheung YB, Thio S, Thumboo J. Using best-worst scaling choice experiments to elicit the most important domains of health for health-related quality of life in Singapore. PLoS One 2018;13(2):e0189687 [FREE Full text] [doi: 10.1371/journal.pone.0189687] [Medline: 29420564]
36. Mühlbacher AC, Kaczynski A, Zweifel P, Johnson FR. Experimental measurement of preferences in health and healthcare using best-worst scaling: an overview. Health Econ Rev 2016;6(1):2. [doi: 10.1186/s13561-015-0079-x] [Medline: 26743636]
37. Cheung KL, Wijnen BFM, Hollin IL, Janssen EM, Bridges JF, Evers SMAA, et al. Using best-worst scaling to investigate preferences in health care. Pharmacoeconomics 2016;34(12):1195-1209. [doi: 10.1007/s40273-016-0429-5] [Medline: 27402349]
38. Sawtooth S. The MaxDiff System Technical Paper (Version 8). Orem, UT: Sawtooth Software, Inc; 2013.
39. Lighthouse Studio Manual. Orem, UT: Sawtooth Software, Inc; 2013.
40. Kongsted A, Nielsen AM. Latent class analysis in health research. J Physiother 2017;63(1):55-58. [doi: 10.1016/j.jphys.2016.05.018] [Medline: 27914733]
41. Abafalvi L, Pénzes M, Urbán R, Foley KL, Kaán R, Kispélyi B, et al. Perceived health effects of vaping among Hungarian adult e-cigarette-only and dual users: a cross-sectional internet survey. BMC Public Health 2019;19(1):302 [FREE Full text] [doi: 10.1186/s12889-019-6629-0] [Medline: 30866897]
42. Benowitz NL, Burbank AD. Cardiovascular toxicity of nicotine: implications for electronic cigarette use. Trends Cardiovasc Med 2016;26(6):515-523. [doi: 10.1016/j.tcm.2016.03.001] [Medline: 27079891]
43. Farsalinos KE, Romagna G, Tsiapras D, Kyrzopoulos S, Voudris V. Characteristics, perceived side effects and benefits of electronic cigarette use: a worldwide survey of more than 19,000 consumers. Int J Environ Res Public Health 2014;11(4):4356-4373 [FREE Full text] [doi: 10.3390/ijerph110404356] [Medline: 24758891]
44. Polosa R, Cibella F, Caponnetto P, Maglia M, Prosperini U, Russo C, et al. Health impact of E-cigarettes: a prospective 3.5-year study of regular daily users who have never smoked. Sci Rep 2017;7(1):13825 [FREE Full text] [doi: 10.1038/s41598-017-14043-2] [Medline: 29150612]
45. Choi S, Lee K, Park SM. Combined associations of changes in noncombustible nicotine or tobacco product and combustible cigarette use habits with subsequent short-term cardiovascular disease risk among South Korean men: a nationwide cohort study. Circulation 2021;144(19):1528-1538 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.121.054967] [Medline: 34601948]
46. Sharman A, Nurmagambetov T. Changes in respiratory function and physical capacity among smokers after switching to iqos: one year follow-up. Glob. J. Respir. Care 2020;6:21-27 [FREE Full text]
47. Polosa R, Morjaria JB, Prosperini U, Busà B, Pennisi A, Malerba M, et al. COPD smokers who switched to e-cigarettes: health outcomes at 5-year follow up. Ther Adv Chronic Dis 2020;11:2040622320961617 [FREE Full text] [doi: 10.1177/2040622320961617] [Medline: 33101622]
48. Lucchiari C, Masiero M, Mazzocco K, Veronesi G, Maisonneuve P, Jemos C, et al. Benefits of e-cigarettes in smoking reduction and in pulmonary health among chronic smokers undergoing a lung cancer screening program at 6 months. Addict Behav 2020;103:106222 [FREE Full text] [doi: 10.1016/j.addbeh.2019.106222] [Medline: 31838445]
49. Laverty AA, Vardavas CI, Filippidis FT. Prevalence and reasons for use of heated tobacco products (HTP) in Europe: an analysis of eurobarometer data in 28 countries. Lancet Reg Health Eur 2021;8:100159 [FREE Full text] [doi: 10.1016/j.lanepe.2021.100159] [Medline: 34557853]
50. Skelton E, Guillaumier A, Tzelepis F, Walsberger S, Paul CL, Dunlop AJ, et al. Alcohol and other drug health-care providers and their client's perceptions of e-cigarette use, safety and harm reduction. Drug Alcohol Rev 2021;40(6):998-1002. [doi: 10.1111/dar.13276] [Medline: 33774886]
51. Xu SS, Meng G, Yan M, Gravely S, Quah ACK, Ouimet J, et al. Reasons for regularly using heated tobacco products among adult current and former smokers in Japan: finding from 2018 ITC Japan survey. Int J Environ Res Public Health 2020;17(21):8030 [FREE Full text] [doi: 10.3390/ijerph17218030] [Medline: 33142757]
52. Betsch C, Böhm R, Airhihenbuwa CO, Butler R, Chapman GB, Haase N, Nurm, et al. Improving medical decision making and health promotion through culture-sensitive health communication: an agenda for science and practice. Med Decis Making 2016;36(7):811-833. [doi: 10.1177/0272989X15600434] [Medline: 26296619]
53. Brick C, McCully SN, Updegraff JA, Ehret PJ, Areguin MA, Sherman DK. Impact of cultural exposure and message framing on oral health behavior: exploring the role of message memory. Med Decis Making 2016;36(7):834-843. [doi: 10.1177/0272989X15570114] [Medline: 25654986]
54. Gerritsen A, Jacobs M, Henselmans I, van Hattum J, Efficace F, Creemers G, et al. Developing a core set of patient-reported outcomes in pancreatic cancer: a delphi survey. Eur J Cancer 2016;57:68-77. [doi: 10.1016/j.ejca.2016.01.001] [Medline: 26886181]

Abbreviations

sf-TNP: smoke-free tobacco and/or nicotine product
TNP: tobacco and/or nicotine product
NRT: nicotine replacement therapy
FDA: Food and Drug Administration

Edited by A Mavragani; submitted 20.03.24; peer-reviewed by N Peraza; comments to author 12.09.24; revised version received 31.10.24; accepted 07.11.24; published 02.01.25.

Please cite as:
Law V, Afolalu EF, Abetz-Webb L, Wemyss LA, Turner A, Chrea C
International Expert Consensus on Relevant Health and Functioning Concepts to Assess in Users of Tobacco and Nicotine Products: Delphi Study
JMIR Form Res 2025;9:e58614
URL: https://formative.jmir.org/2025/1/e58614
doi:10.2196/58614
PMID:

Promoting Digital Health Data Literacy: The Datum Project

Daniel Powell[1*], BCL, JD; Laiba Asad[1*], BCL, JD; Elissa Zavaglia[2], MSc, PhD; Manuela Ferrari[2,3], MHS, PhD

[1]Faculty of Law, McGill University, Montreal, QC, Canada

[2]The Douglas Research Centre, 6875 Boulevard LaSalle (Perry C3 E-3102), Montreal, QC, Canada

[3]Department of Psychiatry, McGill University, Montreal, QC, Canada

[*]these authors contributed equally

Corresponding Author:
Manuela Ferrari, MHS, PhD
The Douglas Research Centre, 6875 Boulevard LaSalle (Perry C3 E-3102), Montreal, QC, Canada

Abstract

With the increased use of digital health innovations in Canadian health care, educating health care users, professionals, and researchers on the ethical challenges and privacy implications of these tools is essential. The Datum project, funded by the Fondation Barreau du Quebec, was created to help these actors better understand legal and ethical issues regarding the collection, use, and disclosure of digital health data for the purposes of scientific research, thereby enhancing literacy around data privacy. The project consists of a multimedia website divided into legislation and policy documents and narrative-based video content. Users can access the core legislation and policies governing the collection and use of health care data geared toward researchers and health practitioners. Users can also view the narrative-based video content explaining key concepts related to digital health data. The Datum project makes an original contribution to the field of law and ethics in health science research by using novel approaches, such as learning health systems and data banks, to improve equity in health care delivery and by generating multimedia content aimed at encouraging health care users to become better consumers and supporting the collective use of their data. The Datum project also promotes digital literacy as a digital communication tool, which has the significant potential to improve health outcomes, bridge the digital divide, and reduce health inequities.

(JMIR Form Res 2025;9:e60832) doi:10.2196/60832

KEYWORDS

health data; digital data; medical records; legislation; ethics; knowledge dissemination; learning health system; data bank

The Digital Health Landscape in Canada

The landscape of digital health data in Canada is multifaceted, encompassing issues of access, cultural relevance, policy implementation, and ethical considerations in using digital health tools in medicine [1]. There is a growing recognition of the potential for digital health innovations to meet the health care needs of Canadians, notwithstanding the challenges and complexities of integrating digital health data in health systems [2]. Public health research highlights the ethical challenges and privacy implications of using digital public health tools, emphasizing the need to find solutions while harnessing their potential [3]. Gasser et al [3] investigated the impact of using big data for health research and surveillance on the public demand for transparency, trust, and fairness. Such ethical challenges may be overcome by balancing risks and benefits to individuals and populations while educating consumers on this topic.

The Datum project aimed to support health care users, professionals, and researchers to better understand the legal and ethical implications of digital health data. There was a need for accessible ways of understanding the collection, storage, and uses of data and how data can be leveraged to meet the objectives of scientific research. Multimedia content and a library of resources were developed to achieve this end.

Background: Genesis of the Datum Project

The word “datum,” the singular of data, originates from the Latin verb *dare*, meaning “to give.” As an act of giving, data may be viewed as not only the result or object of research but also an object arising from a relational context. When patients consent to participate in a research study, they agree to “give” researchers information about their health and health outcomes. The process of collecting, using, and aggregating this information turns it into data that involves the interests of those who originally shared it.

Digital health data are revolutionizing medicine, driving personalized and measurement-based care approaches, informing evidence-based practices, enabling population health management, facilitating early intervention services, supporting research and innovation, and transforming ways of delivering and monitoring health care [4,5]. Analyses of health data reveal the patient’s unique characteristics, risk and protective factors, and treatment responses while providing information on service

outcomes [6]. This allows providers to develop more tailored and precise medical interventions, including targeted therapies and personalized treatment plans while improving the quality of services [7,8].

Influenced by the growing importance of digitalized data, recent legislative and regulatory changes require researchers to carefully consider the relational aspects of data and how data about people is collected, processed, and used. The European Union, through the General Data Protection Regulation, spearheaded early efforts to address these issues, establishing a global standard regarding the responsibilities of those who act as the custodians of data [9]. Recent legislation in the Canadian province of Quebec (eg, Law 25) recognizes the need for enhanced privacy and confidentiality policies, a clearly informed consent process for research participants, and governance frameworks outlining these policies [10].

This rapidly evolving issue calls for enhanced literacy around digital health data, empowering patients, caregivers, and providers to engage with data policies, ask questions, and increase their understanding of their privacy rights. Literacy about data privacy enables individuals to make informed decisions about the extent to which they share their personal information and to advocate for their privacy rights within the broader data governance framework [11].

It is important to empower health care users to support the collective use of data, which extends beyond individual interests, the economic value of data, and commercial use. The current shift toward broader considerations of social and collective value includes introducing democratic forms of governing data production [12]. Moreover, the digital economy should be relational such that individuals are interconnected within population-based data relations [12]. For example, learning health systems align data, technology, and care by continuously aggregating and analyzing ongoing health care data to improve future services, thereby creating a continuous feedback cycle for learning and quality improvement [13,14].

Content of the Datum Website

The Datum project, a multimedia website hosted at McGill University, offers users easy access to core legislation and policy documents that offer an overview of the legislative and policy frameworks governing the collection and use of health care data in Canada and a multimedia section with narrative-based video content capturing realistic exchanges between patients and providers on the collection, use, and sharing of digital health data [15].

Narrative-Based Video Content

Each video animation is set in a real-world context where health care users may be asked to share their digital health data (Figure 1). For example, doctors generally gather intake information from patients seen at medical appointments, including their medical history and symptoms, then upload these data into the patients' electronic health records. Patients who agree to participate in research studies also sign consent forms. Health care users may also identify and share concerns about sharing their personal data through dialogue with an authority figure within the health care system. Their interactions and the responses they receive demonstrate that health care providers appreciate their concerns.

By focusing on the questions of health care users regarding the collection, use, retention, and disclosure of their data in such instances, each video highlights the importance of empowering users to control their data, notably through consent and authorization, data access and transparency, revocation of consent, and data security and privacy. Through these videos, the Datum project hopes to empower health care users to make informed decisions by developing literacy in data privacy and understanding the implications of sharing digital health data in health care and research contexts. The scenarios presented also emphasize the need for collaboration and dialogue between different actors, as well as the advantages of sharing digital health data, such as informing research, empowering future research, and improving health outcomes and the delivery of health care services.

The videos build on each other to provide a comprehensive overview of the legal and ethical implications of using digital health data. The first video, "Introduction to Digital Health Data," focuses on foundational concepts and addresses the difference between data and information. The first and second videos, which situate the characters in both clinical and research contexts, explain some foundational concepts with which health care users may be unfamiliar, such as digital health data and the difference between information and data, as well as primary and secondary uses of digital health data. The third and fourth videos present more complex systems, specifically learning health systems and data banks, where actors in the research and clinical contexts interact with each other and work together to improve health care systems and generate new knowledge. These later videos illustrate the relational aspects of data, specifically how a health care user's data can be used for broader societal purposes. In order to develop learning health systems and data banks which rely on relational databases, health care users have a crucial need to understand the scope and potential of their consent to improve health care.

Figure 1. Image of 4 videos from the Datum multimedia library.

Content Library

As a complement to accessible multimedia content for health care users, the Datum website also features a content library, organized according to legislation and policy documents (Figure 2). This section of the website targets researchers and health care practitioners.

Figure 2. From image of how to select the legislation and policy documents Datum libraries.

Legislation

The regulation of personal information in federalist states like Canada is governed by a complex of provincial and federal rules. The legislation page describes the division of these rules between provincial and federal authorities. At the federal level, the use, collection, and processing of personal information for commercial purposes is subject to the federal Personal Information Protection and Electronic Documents Act. In general, privacy regulation is separated according to rules for private and for public data. Health care data, which may be subject to these privacy regulations, is typically subject to additional rules. The applicable legislation for each province and territory distinguishes three categories of personal information: (1) health information (personal health care information), (2) personal information held by public bodies, and (3) personal information held by private entities.

Policy Documents

The policy documents page provides additional resources for user consultation regarding the collection, use, and disclosure of personal information in the health care context. These materials were selected from the publications of reputable government sources and serve as a starting point for future research.

Feedback and Future Directions

When the Datum website was launched in September 2023, we circulated a survey to gather feedback regarding the content and accessibility of the website. User feedback suggested the utility of incorporating interactive elements into the website to increase user engagement, to allow professionals to contribute to the content library, and as ways of sharing the website with a broader audience.

Ethical Considerations

The Datum project involves the use of fictional characters in the scenarios described in the narrative-based video content. The video content is designed to highlight foundational concepts in the literature and is based on research experience. It does not recommend specific medical procedures or decisions but rather presents interactions between these fictional characters. In addition, usability testing was used to gather feedback to improve the content of the Datum website. Survey responses were not identifiable. No personal information was collected from participants to ensure privacy and confidentiality.

Discussion

Contribution

The Datum project makes an original contribution to the field of law and ethics in health science research. The primary contribution of the project involved the development of literacy related to digital health data through accessible multimedia content and a focused information repository. The presentation of multimedia content transforms traditional privacy practices in the medical sphere into learning opportunities for multiple stakeholders. Digital communication tools, such as mobile health apps, telemedicine, and web-based health information resources have demonstrated a significant potential to improve health and digital literacy, ultimately leading to better health outcomes [16].

Building on the idea that digital communication tools can effectively educate and empower health care users [16], the Datum project particularly seeks to address the data-related concerns of health care users engaged in decision-making processes within the health care context. Users may be unaware of the legislative and institutional measures that protect their digital health data or the various potential uses of their data, whether at the individual or collective levels. Indeed, findings from a survey conducted by Canada Health Infoway, entitled "What Canadians Think: Exploring Public Trust in a Digital Data Ecosystem," revealed that only 3 in 10 respondents had an awareness of Canadian privacy laws protecting their personal health information. Moreover, nearly all participants indicated that, before consenting to the use of their personal health information, they wanted to be informed of how their personal health information would be used and who would have access to it. Surprisingly, a majority of participants expressed readiness to share their personal health information for secondary purposes [17].

Further evidence supporting the promotion of digital literacy in health data, for patients to become better consumers of health care, is rapidly emerging. Informed and activated patients may effectively facilitate positive health outcomes at a lower cost [18], while digital health literacy has also been found to have a significant potential to improve health outcomes, bridge the digital divide, and reduce health inequities [19]. Studies have shown that increasing the numbers of individuals who use professional health websites is an effective measure for improving digital health literacy [20]. Additionally, research has indicated that the better their digital health literacy skills, the more individuals are able to function in the digital environment, exchange data with care providers, and actively participate in the co-design and co-delivery of health services [21]. Our rapidly evolving digital environment underlines the importance of early education in supporting the promotion of digital health literacy as an integral part of school curricula [22]. Finally, research shows that the engagement of health care service users as partners in identifying health research and service improvement priorities has led to the optimization of patient-centered health services [23].

Limitations

Certain limitations in the Datum project should be acknowledged, especially considering the rapid rate of digital transformation. For example, the project content has yet to touch on the impact of using artificial intelligence for digital health data and challenges that the use of artificial intelligence in data modeling may create. As well, the content of the Datum project does not address the concept of data ownership, suggesting future directions for the project.

Conclusions

The evidence from various studies supports the importance of promoting digital literacy concerning health data. The empowerment of health data users is achieved through improved health outcomes, knowledge on how to bridge the digital divide,

active participation in health services, and the optimization of patient-centered care. We hope that the Datum project will empower patients to become better consumers of health care services and research.

Conflicts of Interest

None declared.

References

1. Sekercioglu F, Hamid S. Ontario's digital health vision in the post-COVID-19 pandemic era: a Canadian perspective. J Int Health Sci Manag 2023;9(17):15-22. [doi: 10.48121/jihsam.1132918]
2. Paré G, Leaver C, Bourget C. Diffusion of the digital health self-tracking movement in Canada: results of a national survey. J Med Internet Res 2018 May 2;20(5):e177. [doi: 10.2196/jmir.9388] [Medline: 29720359]
3. Gasser U, Ienca M, Scheibner J, Sleigh J, Vayena E. Digital tools against COVID-19: taxonomy, ethical challenges, and navigation aid. Lancet Digit Health 2020 Aug;2(8):e425-e434. [doi: 10.1016/S2589-7500(20)30137-0] [Medline: 32835200]
4. Abernethy A, Adams L, Barrett M, et al. The promise of digital health: then, now, and the future. NAM Perspect 2022;2022. [doi: 10.31478/202206e] [Medline: 36177208]
5. Walshe K, Rundall TG. Evidence-based management: from theory to practice in health care. Milbank Q 2001;79(3):429-457. [doi: 10.1111/1468-0009.00214] [Medline: 11565163]
6. Ozomaro U, Wahlestedt C, Nemeroff CB. Personalized medicine in psychiatry: problems and promises. BMC Med 2013 May 16;11:132. [doi: 10.1186/1741-7015-11-132] [Medline: 23680237]
7. Hsin H, Fromer M, Peterson B, et al. Transforming psychiatry into data-driven medicine with digital measurement tools. NPJ Digit Med 2018;1:37. [doi: 10.1038/s41746-018-0046-0] [Medline: 31304319]
8. Hulsen T, Jamuar SS, Moody AR, et al. From big data to precision medicine. Front Med (Lausanne) 2019;6:34. [doi: 10.3389/fmed.2019.00034] [Medline: 30881956]
9. Complete guide to GDPR compliance. General Data Protection Regulation (GDPR) EUE. URL: https://gdpr.eu [accessed 2024-12-05]
10. Principaux changements aux lois sur la protection des renseignements personnels. Comission d'accès à l'information (CAI). 2024. URL: https://www.cai.gouv.qc.ca/protection-renseignements-personnels/sujets-et-domaines-dinteret/principaux-changements-loi-25 [accessed 2024-12-05]
11. Wissinger CL. Privacy literacy: from theory to practice. Comm Info Lit 2017;11(2):378-389 [FREE Full text] [doi: 10.15760/comminfolit.2017.11.2.9]
12. Viljoen S. A relational theory of data governance. Yale Law J 2021;131(2):573-654 [FREE Full text]
13. Austin EJ, LeRouge C, Lee JR, et al. A learning health systems approach to integrating electronic patient-reported outcomes across the health care organization. Learn Health Syst 2021 Oct;5(4):e10263. [doi: 10.1002/lrh2.10263] [Medline: 34667879]
14. Ferrari M, Iyer S, LeBlanc A, Roy MA, Abdel-Baki A. A rapid-learning health system to support implementation of early intervention services for psychosis in Quebec, Canada: protocol. JMIR Res Protoc 2022 Jul 19;11(7):e37346. [doi: 10.2196/37346] [Medline: 35852849]
15. Ludic Mind Studio. Datum: developing literacy about digital health data. McGill University. 2024. URL: https://www.mcgill.ca/datum [accessed 2024-12-05]
16. Fitzpatrick PJ. Improving health literacy using the power of digital communications to achieve better health outcomes for patients and practitioners. Front Digit Health 2023;5:1264780. [doi: 10.3389/fdgth.2023.1264780] [Medline: 38046643]
17. Understanding canadians' attitudes, utilization, perceptions and expectations around digital health privacy. Canada Health Infoway. 2024. URL: https://insights.infoway-inforoute.ca/digital-health-privacy-survey [accessed 2024-12-05]
18. Batalden M, Batalden P, Margolis P, et al. Coproduction of healthcare service. BMJ Qual Saf 2016 Jul;25(7):509-517. [doi: 10.1136/bmjqs-2015-004315] [Medline: 26376674]
19. Yang K, Hu Y, Qi H. Digital health literacy: bibliometric analysis. J Med Internet Res 2022 Jul 6;24(7):e35816. [doi: 10.2196/35816] [Medline: 35793141]
20. Liu S, Zhao H, Fu J, et al. Current status and influencing factors of digital health literacy among community-dwelling older adults in Southwest China: a cross-sectional study. BMC Public Health 2022 May 17;22(1):996. [doi: 10.1186/s12889-022-13378-4] [Medline: 35581565]
21. Palumbo R, Nicola C, Adinolfi P. Addressing health literacy in the digital domain: insights from a literature review. K 2022 Dec 19;51(13):82-97. [doi: 10.1108/K-07-2021-0547]
22. Reddy P, Sharma B, Chaudhary K. Digital literacy: A review of literature. International Journal of Technoethics (IJT) 2020 Jul;11(2):65-94. [doi: 10.4018/IJT.20200701.oa1]
23. Murphy F, Noonan MM, Munro D. Conversations and action; combining world cafés with experience-based-co-design to support women to breastfeed. ARCH Women Health Care 2019 Feb 13;2(1):1-4. [doi: 10.31038/AWHC.2019213]

Edited by A Mavragani; submitted 23.05.24; peer-reviewed by B Sterner; revised version received 21.10.24; accepted 28.10.24; published 03.01.25.

Please cite as:
Powell D, Asad L, Zavaglia E, Ferrari M
Promoting Digital Health Data Literacy: The Datum Project
JMIR Form Res 2025;9:e60832
URL: https://formative.jmir.org/2025/1/e60832
doi:10.2196/60832

Quantifying the Regional Disproportionality of COVID-19 Spread: Modeling Study

Kenji Sasaki[1], PhD; Yoichi Ikeda[1], PhD; Takashi Nakano[1,2], PhD

[1]Center for Infectious Disease Education and Research, Osaka University, Co-creation BLDG. D88-1, 2-1 Yamadaoka, Suita, Osaka, Japan
[2]Research Center for Nuclear Physics, Osaka University, Ibaraki, Osaka, Japan

Corresponding Author:
Kenji Sasaki, PhD
Center for Infectious Disease Education and Research, Osaka University, Co-creation BLDG. D88-1, 2-1 Yamadaoka, Suita, Osaka, Japan

Abstract

Background: The COVID-19 pandemic has caused serious health, economic, and social consequences worldwide. Understanding how infectious diseases spread can help mitigate these impacts. The Theil index, a measure of inequality rooted in information theory, is useful for identifying geographic disproportionality in COVID-19 incidence across regions.

Objective: This study focused on capturing the degrees of regional disproportionality in incidence rates of infectious diseases over time. Using the Theil index, we aim to assess regional disproportionality in the spread of COVID-19 and detect epicenters where the number of infected individuals was disproportionately concentrated.

Methods: To quantify the degree of disproportionality in the incidence rates, we applied the Theil index to the publicly available data of daily confirmed COVID-19 cases in the United States over a 1100-day period. This index measures relative disproportionality by comparing daily regional case distributions with population proportions, thereby identifying regions where infections are disproportionately concentrated.

Results: Our analysis revealed a dynamic pattern of regional disproportionality in the confirmed cases by monitoring variations in regional contributions to the Theil index as the pandemic progressed. Over time, the index reflected a transition from localized outbreaks to widespread transmission, with high values corresponding to concentrated cases in some regions. We also found that the peaks in the Theil index often preceded surges in confirmed cases, suggesting its potential utility as an early warning signal.

Conclusions: This study demonstrated that the Theil index is one of the effective indices for quantifying regional disproportionality in COVID-19 incidence rates. Although the Theil index alone cannot fully capture all aspects of pandemic dynamics, it serves as a valuable tool when used alongside other indicators such as infection and hospitalization rates. This approach allows policy makers to monitor regional disproportionality efficiently, offering insights for early intervention and targeted resource allocation.

(JMIR Form Res 2025;9:e59230) doi:10.2196/59230

KEYWORDS

infectious disease; COVID-19; epidemiology; public health; SARS-CoV-2; pandemic; inequality measure; information theory; Kullback-Leibler divergence

Introduction

The COVID-19 pandemic has caused serious health problems and has had major economic and social consequences worldwide. It has highlighted the need to understand regional disparities in infection rates to strengthen public health responses since infection dynamics are influenced by factors such as population density, socioeconomic conditions, and health care infrastructure [1,2]. Numerous indicators and models have been proposed to address the problem, and mechanisms for the spread of the infection and intervention measures to control the pandemic have been studied [3-9].

Several recent studies have investigated regional differences in COVID-19 prevalence [10-13]. Differences in the prevalence rates between regions highlight the need to understand regional inequalities in pandemic response strategies. Effectively addressing these disparities requires accurate quantification and understanding of regional disproportionalities in daily confirmed COVID-19 cases.

In the field of economics, various indicators have been developed to measure resource and income inequality, including an index proposed by Theil, which incorporated information theory [14]. Manz and Mansmann [15] have demonstrated the importance of using inequality indices for monitoring changes in geographic inequality; for instance, the Theil index was used to track geographic disproportionality over time during the COVID-19 pandemic, providing important insights for public health policy.

The aim of this paper is to quantify the interregional disproportionality in the number of confirmed cases using the Theil index, which mathematically corresponds to the Kullback-Leibler (KL) divergence in information theory [16]. The Theil index is an effective method of measuring the degree of disproportionality and objectively assessing biases in the interregional distribution of infected individuals.

Methods

Overview

We analyzed the time trends of daily COVID-19–confirmed cases in the United States over 1100 days since the first reported case on January 21, 2020 [17]. Data are taken from the COVID-19 data repository at the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University [18]. US state population data were obtained from the US Census Bureau website [19]. Population changes due to migration, births, and deaths were not considered in the analysis.

The Disproportionality Measure: Theil Index

The Theil index is commonly applied in various fields including economics, sociology, and information theory. It quantifies the relative differences between various components of a dataset. In the context of regional analysis of the confirmed cases, the Theil index can be employed to evaluate the distribution of infected individuals across different regions. In this study, we use the Theil index to identify regions with disproportionate numbers of confirmed cases relative to their population size.

The discrete form of the Theil index is expressed as:

$$T=\sum_{i=1}^{N} t_i=\sum_{i=1}^{N} p_i \ln \frac{p_i}{q_i}$$

where N is the total number of regions being considered and ln is the natural logarithm. The Theil index T is composed of a sum of t_i which is a partial contribution from region i. The discrete probability distribution, p_i, in region i is defined as the ratio of daily confirmed cases region i to the total confirmed cases across all regions for that day. Similarly, the population ratio, q_i, in region i is the ratio of the population in region i to the total population across all regions.

The Theil index, which is mathematically related to the KL divergence, is a nonsymmetric metric that measures the relative entropy or informational difference between two distributions. It is sensitive to the interregional variations in the distribution of the confirmed cases, with its maximum value attained when the confirmed cases are concentrated in areas with the smallest population proportion. Consequently, the index tends to exhibit higher values when a small number of regions account for a large share of the confirmed cases, and conversely, lower values when the confirmed cases are more evenly distributed across regions. Notably, it remains nonnegative and reaches a minimum value of 0 only when the two distributions are identical. Therefore, applying the Theil index to the time-series data of the confirmed cases, and monitoring changes in the index over time, we quantified the degree of spread of COVID-19 cases and assessed whether the confirmed cases were disproportionately concentrated relative to the regional population sizes over time.

Ethical Considerations

This study used publicly available, deidentified COVID-19 data from CSSE at Johns Hopkins University [18], and therefore, additional ethics approval and informed consent were not required. The aggregated data ensured privacy and confidentiality, and no direct human participants were involved; thus, no compensation was provided. No identifiable information appears in any images or materials.

Results

To address fluctuations in the Theil index caused by data aggregation inconsistencies during holidays across different regions, the 7-day average of confirmed COVID-19 cases was used instead of raw data.

Figure 1 illustrates a 2-axis graph showing the time trends of the Theil index (left axis) and the number of confirmed cases in a logarithmic scale (right axis, logarithmic scale). The horizontal axis represents the number of days elapsed (denoted by d in the text) since the date of the first reported case in the United States.

In Figure 1, there are eight notable surges of the confirmed cases, occurring at approximately d=80 (first), 180 (second), 350 (third), 450 (fourth), 580 (fifth), 720 (sixth), 900 (seventh), and 1080 (eighth), respectively. The presence of multiple peaks in the Theil index indicates that infected individuals were concentrated in specific regions during the period, and the degree of this concentration can be assessed by examining the numerical values. However, it is important to note that changes in the Theil index simply indicate the degree of regional disproportionality in the confirmed cases rather than absolute increases or decreases in the number of infected individuals. Therefore, this indicator is most effective when interpreted in conjunction with actual trends in the number of confirmed cases.

Before the first peak, the number of confirmed cases was quite low, and the Theil index fluctuated erratically. As d increased near the first peak, the Theil index gradually decreased, reaching a local minimum around d=120. This suggested that the initially localized epidemic began to spread throughout the US during the early stages of the global pandemic. Similar trends were observed during subsequent surges, such as slight a increase in the Theil index before the peak, followed by a decrease. This could be seen as a precursor to a surge in the number of infected individuals. This finding aligns with previous research by Ikeda, Sasaki, and Nakano [7].

The following examples provide interesting insights; when the Theil index value was high and the number of confirmed cases was low (d=60, 550, etc), it indicated that the infectious disease was localized and beginning to spread to various regions. Conversely, when the index was low and the number of confirmed cases was high (d=750, etc), it indicated that there was no obvious epicenter of the infectious disease, with the number of confirmed cases increasing relatively and evenly across different regions.

The contributions to the Theil index from each region (t_i), calculated from the number of cases on each date, were arranged

in chronological order and visualized using a heatmap, as shown in Figure 2. Regions with a high proportion of confirmed cases are represented in red, while regions with a low proportion are colored blue. Notably, there are long intervals between the deep red patches in some regions such as California, Florida, and New York. Particularly, the periods of intense infection represented by these deep red patches were not repeated at short intervals. This phenomenon is of great importance in infectious disease management. Once a major epidemic in an area has subsided, the interval between subsequent outbreaks provides an opportunity to rebuild the health care infrastructure and implement preventive measures before the occurrence of the next epidemic.

Based on the observations from Figure 2, the epicenter of infectious diseases as indicated by the red patches alternates between New York, California, and Florida. This insight is crucial for understanding the underlying mechanisms of the spread of infectious diseases in the future. Furthermore, after d=750, both the red and blue colors fade over time, indicating the absence of a single epicenter, and a widespread outbreak of COVID-19. This pattern suggests the ineffectiveness of countermeasures against the spread of infectious diseases under these circumstances.

Figure 3A shows the contributions to the Theil index by region at d=60. The horizontal axis in the figure shows the state code (as listed in Multimedia Appendix 1). There is a significant contribution to the Theil index from New York State compared to the other regions. Figure 3B shows that at this point confirmed cases were highly localized in these regions.

There were relatively large negative contributions to the Theil index from California, Florida, and Texas, which were regions with high population ratios. It is interesting to note that there was little risk of infection in these regions at that point; however, the number of infected individuals rapidly increased following the concentration of confirmed cases in New York.

Figure 4 shows the contributions to the Theil index from each region at d=550 and 750. At d=550 shown in Figure 4A, the Theil index reaches a peak, and the trend of confirmed cases is increasing. This suggests that a new epidemic is emerging, mainly in Florida and Louisiana. However, their contributions are significantly smaller compared to New York at d=60, as seen in Figure 3. This indicates that regional disproportionality is much less pronounced than in the early stage of the COVID-19 pandemic. It is also interesting to look at data on d=750 as shown in Figure 4B, when confirmed cases in the United States are at their maximum. Although several regions show large contributions to the Theil index, the epicenter of COVID-19 is no longer obvious, suggesting that COVID-19 cases are uniformly distributed across the country.

Figure 1. Time trends of the Theil index on the left axis and the 7-day average number of the confirmed cases on the right axis on a logarithmic scale are shown in the red and blue curves, respectively. The horizontal axis is the number of days elapsed since January 21, 2020.

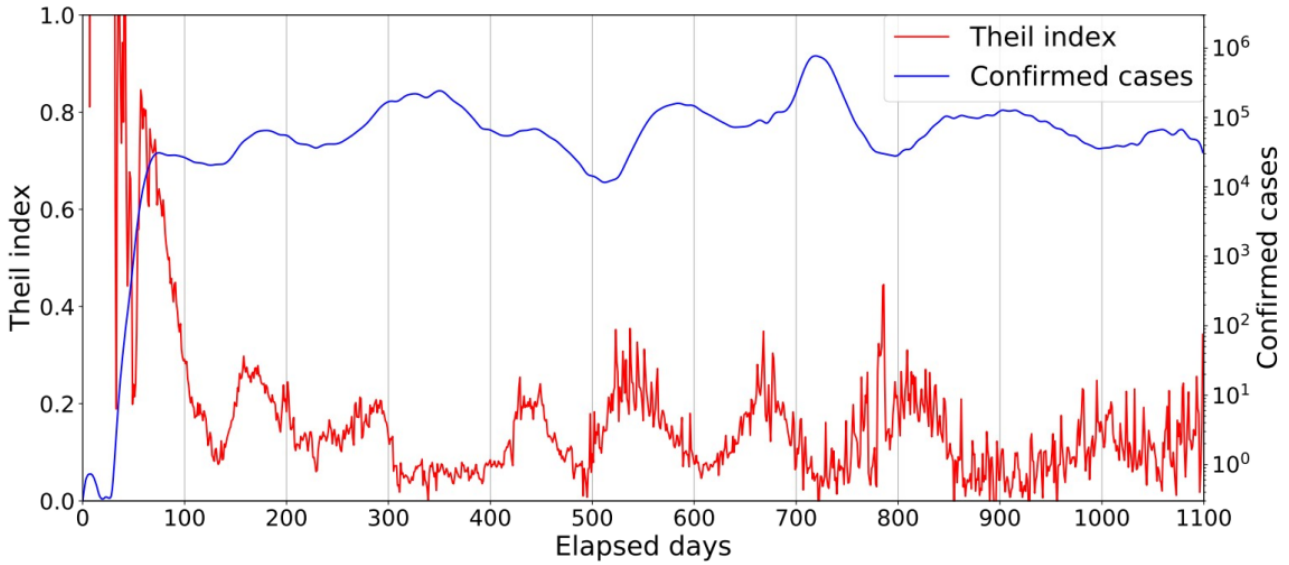


Figure 2. Partial contributions to the Theil index from each region, t_i, are displayed in a heatmap over time. The horizontal axis represents the number of days elapsed since January 21, 2020. The vertical axis shows the names of states in the United States. The positive (high concentration of incidences) and negative (low concentration) contributions to the Theil index correspond to deep red and blue colors, respectively.

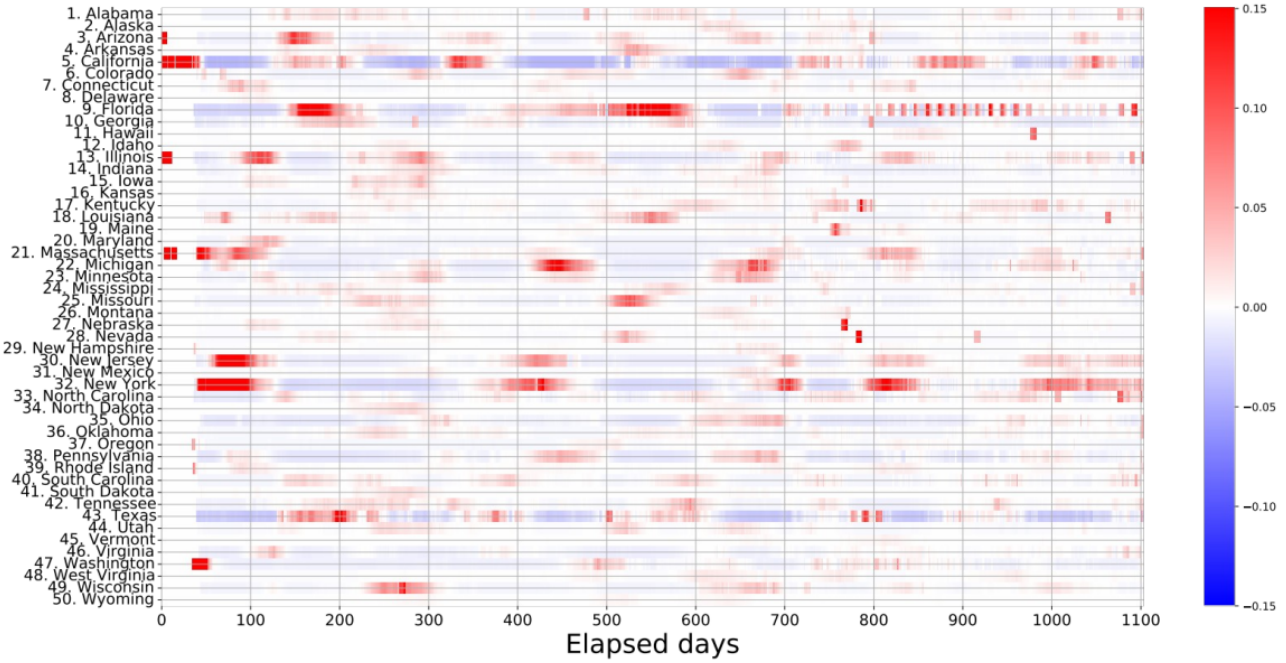


Figure 3. Partial contributions to the Theil index from each region, t_i, at d=60. The horizontal axis shows the state code given in Multimedia Appendix 1. (A) Comparison of the distribution of the confirmed cases and population. The vertical axis shows the ratio of a part to the whole region for populations and for the confirmed cases. (B) Contributions to the Theil index from each region. The vertical axis shows the strength of the contribution to the Theil index. The significantly high value of partial contribution to the Theil index is highlighted in orange.

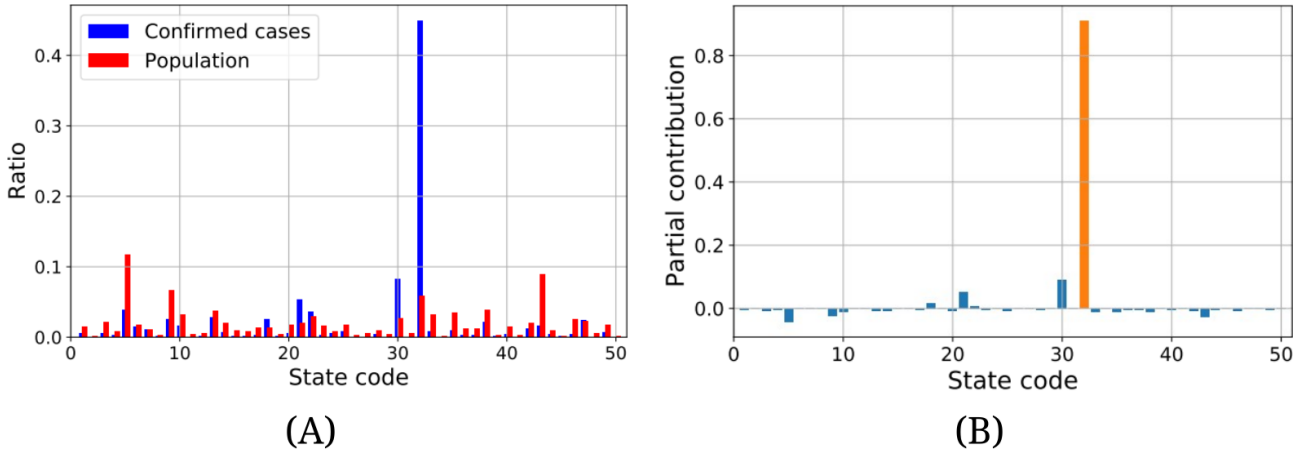
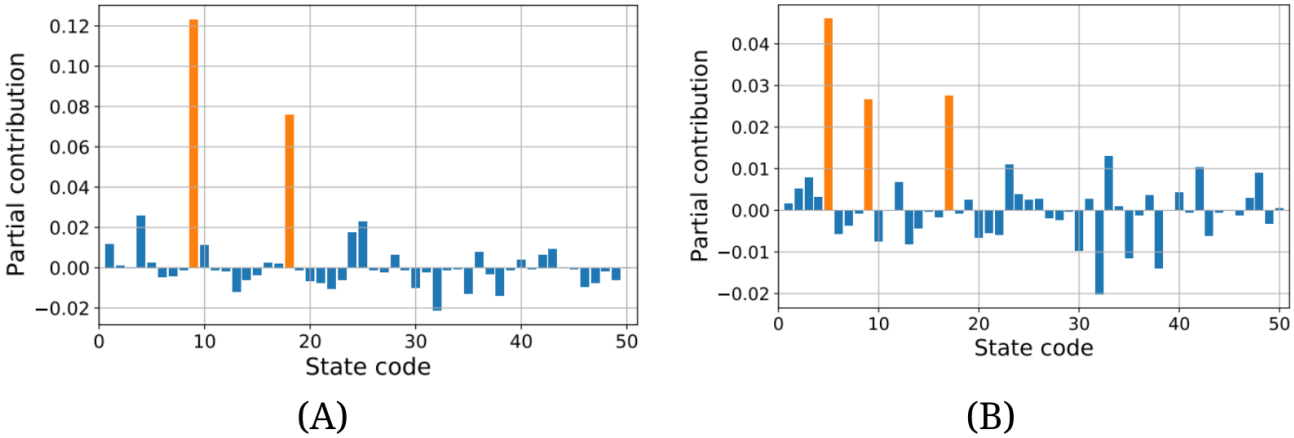


Figure 4. Partial contributions of the Theil index from each region, t_i at a specific date. The vertical axis shows the strength of contribution to the Theil index. The horizontal axis shows the state code given in Multimedia Appendix 1. (A) Contributions of the Theil index at d=550. (B) Contributions of the Theil index at d=750. The significantly high values of partial contribution to the Theil index are highlighted in orange.



Discussion

This study demonstrates the utility of the Theil index for quantifying regional disproportionalities in the distribution of COVID-19 cases. It offers an intuitive and efficient approach to identifying hotspots and monitoring the spread of infection. However, certain limitations may affect result interpretation.

The accuracy of the analysis depends on data quality; factors such as underreporting, delays in case confirmation, and regional differences in testing capacity may introduce biases into case counts. These issues could potentially impact the calculated Theil index and the assessment of regional disproportionalities.

Additionally, this study focuses primarily on confirmed cases rather than new infections, limiting its capacity to predict future spread. Therefore, the Theil index alone may not be sufficient for determining the timing and location of public health interventions, such as isolation measures. To support comprehensive policy-making, it should be used alongside other indicators, such as infection rates, hospitalization rates, and health care capacity.

Conventional spatiotemporal analysis methods [20,21] are widely used in epidemiology and public health to track infectious disease spread and visualize infection clusters over time in specific regions. These established tools effectively detect geographical clusters, identify areas with unusually high incidence, and reveal disease hotspots within defined spatial ranges.

In contrast, our method offers two distinct advantages. First, an increase in the Theil index acts as a precursor to a surge in the number of infected individuals. Second, it quantifies regional disproportionalities in incidence rates at any given time. Unlike conventional methods that emphasize physical distance and spatial proximity, our approach treats regions as discrete categories to calculate incidence rate disproportionalities. Although simple, this approach provides an intuitive way to identify epicenters at a lower computational cost compared to spatiotemporal scanning, enabling us to detect early surges in confirmed cases and pinpoint regions with concentrated infections.

For instance, the concentration of COVID-19 cases in New York at d=60 as shown in Figure 3A and B, cannot be overlooked when considering infection control. The lockdown was implemented in New York City [22] and coincided with a period when the contribution to the Theil index was concentrated in New York State. Although it is challenging to assess the direct impact of lockdown using the Theil index alone, the timing appears appropriate based on the pattern of concentration of confirmed cases.

Integrating our method with additional data sources, such as mobility patterns and health care capacity, will enhance pandemic response strategies, particularly for early intervention and efficient resource allocation.

In conclusion, this study demonstrates the application of the Theil index in quantifying regional disproportionalities in confirmed cases and monitoring their evolution over time. By analyzing confirmed case data in the United States, we have identified patterns of disproportionalities, specified epicenters, and characterized localized outbreaks.

Continued monitoring and analysis of regional differences in COVID-19 transmission remain essential, especially considering emerging variants and evolving public health responses. Our findings highlight the importance of understanding the regional dynamics of infected individuals for effective pandemic response interventions.

Incorporating the findings of this study will help policy makers refine strategies and address the diverse needs of different regions, ultimately increasing the effectiveness of pandemic response efforts and mitigating the impact of future health crises.

Lastly, the decomposability of the Theil index makes it possible to quantify and compare disproportionality in groups with specific characteristics, such as age, vaccination coverage, and health care accessibility. Identifying these disproportionalities will provide important insights for future pandemic responses.

Acknowledgments

We thank the members of the Division of Scientific Information and Public Policy (SiPP) at the Center for Infectious Disease Education and Research (CiDER) Osaka University for useful discussions. This research was supported by The Nippon Foundation—Osaka University Project for Infectious Disease Prevention.

Data Availability

Data were derived from public resources.

Conflicts of Interest

None declared.

Multimedia Appendix 1
US states codes.
[PNG File, 58 KB - formative_v9i1e59230_app1.png]

References

1. Lopez L, Hart LH, Katz MH. Racial and ethnic health disparities related to COVID-19. JAMA 2021 Feb 23;325(8):719-720. [doi: 10.1001/jama.2020.26443] [Medline: 33480972]
2. Adhikari S, Pantaleo NP, Feldman JM, Ogedegbe O, Thorpe L, Troxel AB. Assessment of community-level disparities in coronavirus disease 2019 (COVID-19) infections and deaths in large US metropolitan areas. JAMA Netw Open 2020 Jul 1;3(7):e2016938. [doi: 10.1001/jamanetworkopen.2020.16938] [Medline: 32721027]
3. Perkins TA, España G. Optimal control of the COVID-19 pandemic with non-pharmaceutical interventions. Bull Math Biol 2020 Sep 4;82(9):118. [doi: 10.1007/s11538-020-00795-y] [Medline: 32888118]
4. Nakano T, Ikeda Y. Novel indicator to ascertain the status and trend of COVID-19 spread: modeling study. J Med Internet Res 2020 Nov 30;22(11):e20144. [doi: 10.2196/20144] [Medline: 33180742]
5. Buchy P, Buisson Y, Cintra O, et al. COVID-19 pandemic: lessons learned from more than a century of pandemics and current vaccine development for pandemic control. Int J Infect Dis 2021 Nov;112:300-317. [doi: 10.1016/j.ijid.2021.09.045] [Medline: 34563707]
6. Stenseth NC, Dharmarajan G, Li R, Shi ZL, Yang R, Gao GF. Lessons learnt from the COVID-19 pandemic. Front Public Health 2021;9:694705. [doi: 10.3389/fpubh.2021.694705] [Medline: 34409008]
7. Ikeda Y, Sasaki K, Nakano T. A new compartment model of COVID-19 transmission: the broken-link model. Int J Environ Res Public Health 2022 Jun 3;19(11):6864. [doi: 10.3390/ijerph19116864] [Medline: 35682447]
8. Pandemic preparedness. Nature. 2022 Oct 26. URL: https://www.nature.com/collections/jaacfgeief [accessed 2024-03-14]
9. Sasaki K, Ikeda Y, Nakano T. The effects of behavioral restrictions on the spread of COVID-19. Rep 2022;5(4):37. [doi: 10.3390/reports5040037]
10. Jinjarak Y, Ahmed R, Nair-Desai S, Xin W, Aizenman J. Accounting for global COVID-19 diffusion patterns, January-April 2020. Econ Disaster Clim Chang 2020;4(3):515-559. [doi: 10.1007/s41885-020-00071-2] [Medline: 32901228]
11. Mollalo A, Vahedi B, Rivera KM. GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. Sci Total Environ 2020 Aug 1;728:138884. [doi: 10.1016/j.scitotenv.2020.138884] [Medline: 32335404]
12. Villanustre F, Chala A, Dev R, et al. Modeling and tracking Covid-19 cases using big data analytics on HPCC system platform. J Big Data 2021;8(1):33. [doi: 10.1186/s40537-021-00423-z] [Medline: 33614394]
13. Yue H, Hu T. Geographical detector-based spatial modeling of the COVID-19 mortality rate in the continental United States. Int J Environ Res Public Health 2021 Jun 25;18(13):6832. [doi: 10.3390/ijerph18136832] [Medline: 34202168]
14. Theil H. Economics and Information Theory: North-Holland Publishing Company: Amsterdam, North-Holland; 1967.
15. Manz KM, Mansmann U. Inequality indices to monitor geographic differences in incidence, mortality and fatality rates over time during the COVID-19 pandemic. PLoS ONE 2021;16(5):e0251366. [doi: 10.1371/journal.pone.0251366] [Medline: 33984055]
16. Kullback S, Leibler RA. On information and sufficiency. Ann Math Statist 1951 Mar;22(1):79-86. [doi: 10.1214/aoms/1177729694]
17. Holshue ML, DeBolt C, Lindquist S, et al. First case of 2019 novel coronavirus in the United States. N Engl J Med 2020 Mar 5;382(10):929-936. [doi: 10.1056/NEJMoa2001191] [Medline: 32004427]
18. COVID-19 data repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. Github. URL: https://github.com/CSSEGISandData/COVID-19 [accessed 2024-03-14]
19. National population totals and components of change: 2020-2023. US Census Bureau. URL: https://www.census.gov/data/datasets/time-series/demo/popest/2020s-national-total.html [accessed 2024-03-14]
20. Kulldorff M, Heffernan R, Hartman J, Assunção R, Mostashari F. A space-time permutation scan statistic for disease outbreak detection. PLoS Med 2005 Mar;2(3):e59. [doi: 10.1371/journal.pmed.0020059] [Medline: 15719066]

21. Desjardins MR, Hohl A, Delmelle EM. Rapid surveillance of COVID-19 in the United States using a prospective space-time scan statistic: detecting and evaluating emerging clusters. Appl Geogr 2020 May;118:102202. [doi: 10.1016/j.apgeog.2020.102202] [Medline: 32287518]
22. Huang Y, Li R. The lockdown, mobility, and spatial health disparities in COVID-19 pandemic: a case study of New York City. Cities 2022 Mar;122:103549. [doi: 10.1016/j.cities.2021.103549] [Medline: 35125596]

Abbreviations

CSSE: Center for Systems Science and Engineering
KL: Kullback-Leibler

Edited by A Mavragani; submitted 06.04.24; peer-reviewed by A Inostrosa-Psijas, Z Guo; revised version received 04.11.24; accepted 06.11.24; published 03.01.25.

Please cite as:
Sasaki K, Ikeda Y, Nakano T
Quantifying the Regional Disproportionality of COVID-19 Spread: Modeling Study
JMIR Form Res 2025;9:e59230
URL: https://formative.jmir.org/2025/1/e59230
doi:10.2196/59230

Original Paper

Quality Assessment of Medical Institutions' Websites Regarding Prescription Drug Misuse of Glucagon-Like Peptide-1 Receptor Agonists by Off-Label Use for Weight Loss: Website Evaluation Study

Rie Oyama[1], MD; Tsuyoshi Okuhara[1], MPH, PhD; Emi Furukawa[2], MD, PhD; Hiroko Okada[1], MPH, PhD; Takahiro Kiuchi[1], MD, PhD

[1]Department of Health Communication, School of Public Health, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan
[2]University Hospital Medical Information Network, The University of Tokyo Hospital, Tokyo, Japan

Corresponding Author:
Rie Oyama, MD
Department of Health Communication
School of Public Health
Graduate School of Medicine, The University of Tokyo
7-3-1 Hongo
Bunkyo-ku
Tokyo, 113-8655
Japan
Phone: 81 3 5800 6549
Fax: 81 3 5689 0726
Email: rooyama-oss@umin.ac.jp

Abstract

Background: Misuse of glucagon-like peptide-1 receptor agonists (GLP-1RAs) has emerged globally as individuals increasingly use these drugs for weight loss because of unrealistic and attractive body images advertised and shared on the internet.

Objective: This study assesses the quality of information and compliance with Japan's medical advertising guidelines on the websites of medical institutions that prescribe GLP-1RAs off-label for weight loss.

Methods: Websites were identified by searching Google and Yahoo! by using keywords related to GLP-1RAs and weight loss in August 2024. The quality of information on these websites was assessed using the DISCERN instrument. To comply with Japan's medical advertising guidelines, we evaluated whether the 5 mandatory items for advertisements of self-paid medical treatments involving the off-label use of drugs were stated and whether there were any exaggerated claims. The content of the exaggerated advertisements was categorized into themes.

Results: Of the 87 websites included, only 1 website stated all 5 mandatory items. Websites listing "ineligible for the relief system for sufferers from adverse drug reactions" had the lowest percentage at 9% (8/87), while 83% (72/87) of the websites listed exaggerated advertisements. Approximately 69% (60/87) of the websites suggested that no exercise or dietary therapy was required, 24% (21/87) suggested that using GLP-1RAs is a natural and healthy method, and 31% (27/87) of the websites provided the author's personal opinions on the risks of using GLP-1RAs. The mean total DISCERN score for all 87 websites was 32.6 (SD 5.5), indicating low quality. Only 1 website achieved a good rating, and 9 websites were rated as fair. The majority of the websites were rated as poor (72 websites) or very poor (5 websites).

Conclusions: We found that the quality of information provided by the websites of medical institutions prescribing GLP-1RAs off-label for weight loss was very low and that many websites violated Japan's medical advertising guidelines. The prevalence of exaggerated advertisements, which may lead consumers to believe that they can lose weight without dietary or exercise therapy, suggests the risk of GLP-1RA misuse among consumers. Public institutions and health care providers should monitor and regulate advertisements that violate guidelines and provide accurate information regarding GLP-1RAs, obesity, and weight loss.

(JMIR Form Res 2025;9:e68792) doi:10.2196/68792

KEYWORDS

prescription drug misuse; GLP-1 receptor agonists; off-label use; weight loss; information quality; DISCERN; web-based information; information provision; misinformation; advertising guidelines; exaggerated advertisements

Introduction

Prescription drug misuse is an emerging global public health issue [1-3]. According to a previous study [4], misuse is defined as the therapeutic use of legal drugs in a manner other than as prescribed, and abuse is defined as the use of illegal drugs for nontherapeutic purposes such as psychological pleasure and weight loss. The misuse and abuse of drugs can lead to health problems. In the United States and European countries, misuse of opioid analgesics, benzodiazepines, over-the-counter cough suppressants containing dextromethorphan and codeine, and psychostimulants have been observed [1-3,5]. Further, in Japan, misuse of over-the-counter cough suppressants, pain relievers, antianxiety medications, and sedative drugs has been observed [6,7]. However, misuse of glucagon-like peptide-1 receptor agonists (GLP-1RAs) has become a problem in recent years [8-10].

GLP-1RAs are the most effective drugs for obesity treatment [11]. Semaglutide (Wegovy), a type of GLP-1RA, was approved by the US Food and Drug Administration (FDA) as an obesity treatment drug in 2021 [12]. Since then, its market share has rapidly expanded [13]. Conversely, multiple news media outlets have reported concerns about the misuse of GLP-1RAs [14,15]. Due to unrealistic and attractive body images that are advertised by celebrities on social media, people without obesity allegedly obtain and use GLP-1RAs illicitly for weight loss purposes [16]. Adverse events of GLP-1RA misuse have also been reported [8,9]. A previous study showed that the use of semaglutide in patients with overweight/obesity may be associated with the incidence of nonarteritic anterior ischemic optic neuropathy, a severe ophthalmologic condition that can lead to blindness [17]. Another previous study showed that GLP-1RA users for weight management had an increased risk of serious gastrointestinal disorders compared to bupropion-naltrexone users, with adjusted hazard ratios for pancreatitis, bowel obstruction, and gastroparesis reported as 9.09, 4.22, and 3.67, respectively [18]. However, the use of GLP-1RAs for weight management is a relatively new therapeutic approach, and the long-term side effects associated with their use remain unknown. Additionally, the effectiveness and potential adverse effects of using GLP-1RAs for further weight loss in individuals without obesity also remain unknown. This increased demand has caused a global shortage of GLP-1RAs [16]. The European Medicines Agency has announced recommendations by manufacturers, health care professionals, and citizens for the proper use of GLP-1RAs [19].

In Japan, Wegovy was approved for the treatment of obesity on March 27, 2023. According to the Optimal Use Promotion Guideline by the Japan's Ministry of Health, Labor, and Welfare, patients eligible for obesity treatment are "those with BMI ≥35 kg/m² or those with BMI ≥27 kg/m² who also have 2 or more health conditions related to obesity and who do not achieve sufficient results from diet and exercise therapy" [20]. The guidelines also strictly prohibit the use of GLP-1RAs for weight loss purposes without obesity as well as institutional criteria for prescribing GLP-1RAs under insurance-covered medical care [20,21]. Japan's Ministry of Health, Labor, and Welfare released a "Request for cooperation due to tight inventory of GLP-1 receptor agonist" [22], and the Japan Diabetes Society [23] and Japan Society for the Study of Obesity [24] have made recommendations regarding the off-label use of GLP-1RAs.

However, several private clinics in Japan prescribe GLP-1RAs for off-label weight loss. Japan's National Consumer Affairs Center and Consumer Affairs Agency warns citizens against the use of GLP-1RAs for weight loss purposes because of the possibility of medical institutions prescribing GLP-1RAs through advertisements that are prohibited by guidelines, inadequate physician consultations and explanations, and inadequate systems to manage adverse events [25,26]. Websites of medical institutions in Japan are regulated by the Medical Care Act [27] and the Ministry of Health, Labor and Welfare's "Regulations on Advertisements of Hospitals, etc under the Medical Care Act" (Japan's medical advertising guidelines) [28], which prohibit false or exaggerated advertisements. The reason for this is "because false advertisements may cause patients to lose the opportunity to receive appropriate medical consultation or receive inappropriate medical treatments by giving them information that is significantly different from the facts, etc"[28].

In previous studies on health information quality, these issues were assessed as part of the reliability of health information. The DISCERN instrument developed by Charnock et al [29] has been used in many studies to assess the publication reliability and the quality of the written health information (Multimedia Appendix 1). The validity and reliability of the DISCERN instrument were ensured, and in addition to the original English language version, Chinese [30], Spanish [31], Brazilian Portuguese [32], and Japanese [33] versions have been developed [29]. Previous studies using the DISCERN instrument have examined the quality of information on websites regarding posttraumatic stress disorder [34], breast cancer [35], nasopharyngeal carcinoma [36], and cosmetic injectability [37]. These studies showed that the quality of individual websites varied considerably: the quality of information was higher for academic and government institutions that operate websites and lower for commercial institutions [34-37]. A systematic review showed that in research using the DISCERN instrument, none found the mean DISCERN score of all websites to be excellent, and the information quality of the majority of websites ranged from fair to very good [38].

In recent years, most people have searched the internet for health information [39]. The internet usage rate for individuals in Japan in 2022 was 84.9% and exceeded 90% when limited to the 10-50 years age range [40]. For cosmetic treatments, 95% of the patients search for information on the internet before entering a doctor's office [41]. Hence, if medical institutions' websites

on GLP-1RAs contain misleading, exaggerated, or unreliable information, these websites should be immediately assessed and improved. However, to our knowledge, no study has investigated the quality of information that people without obesity obtain from websites when they consider obtaining GLP-1RAs for weight loss. This study aims to assess compliance with medical advertising guidelines and the quality of information on the websites of medical institutions that prescribe GLP-1RAs off-label for weight loss in Japan.

Methods

Data Collection

Search Terms

We used a Japanese-language search string input into Google Japan [42] and Yahoo Japan [43], which has the largest search engine market shares in Japan, in August 2024 [44]. The GLP-1RAs approved by the FDA for obesity treatment are liraglutide, semaglutide, and tirzepatide that are sold in Japan under the product names Rybelsus, Ozempic, Victoza, Saxenda, and Mounjaro. The search keywords comprised 6 in total (Rybelsus diet, Ozempic diet, Victoza diet, Saxenda diet, Mounjaro diet, and GLP-1 diet), which included each product name combined with diet (used specifically to mean weight loss in Japan), as well as the comprehensive term GLP-1 diet. Wegovy was excluded from the search terms because it was the newest drug and was not available in the general market in Japan. Saxenda is approved in Japan for the treatment of diabetes under the name Victoza, but private clinics that use the drug off-label advertise it under the name Saxenda; so, it was included in the search terms.

All words were entered in the search window, one keyword at a time, and 20 websites were reviewed per search engine for each term. This is because a previous study showed that 97.2% of the users only view the first 1-10 websites of search results, and 2% click on links on the second page of the search results and beyond [45]. Google Chrome's incognito browser that would have all history deleted was used, and personal accounts were not used.

Inclusion and Exclusion Criteria

The inclusion criteria were as follows: websites written in Japanese; websites created by medical institutions (with medical department labels); websites clearly stating that GLP-1RAs were being sold; and the medical institutions operating the websites were not accredited as facilities capable of treating obesity using GLP-1RAs by the Japanese Circulation Society, Japan Diabetes Society, or Japan Endocrine Society. The exclusion criteria were as follows: duplicate websites, government or academic statements, PDF documents, news reports, affiliate advertisements, and websites that were unclear whether they prescribe GLP-1RAs. The first author (RO) reviewed all the websites in the search results and extracted eligible websites.

Data Extraction

RO recorded the names and URLs of all the websites, names of the medical institutions operating the websites, the medical departments they represented, and the types of GLP-1RAs they sold. If not listed, it was recorded as unknown. Medical institutions can represent multiple medical departments in Japan. Therefore, if several departments were listed, all were recorded. All data were recorded in a Microsoft Excel 2021 spreadsheet by RO.

Data Assessment

First, all identified websites were assessed by an internal medicine physician (RO) for compliance with the medical advertising guidelines in Japan and quality by using the DISCERN instrument. Second, the third author (EF, an internal physician) was trained through a 1-hour meeting and a 1-hour pilot assessment, and then EF independently assessed approximately 20% (20/87) of the total websites. These 20 websites were randomly selected using Microsoft Excel to generate random numbers. Finally, they assessed the agreement between the coders of the 20 websites. The results were recorded using a Microsoft Excel 2021 spreadsheet. More precise information is provided below.

Assessment of Compliance With Medical Advertising Guidelines

In the case of noninsured medical treatments involving off-label use of drugs, it is prohibited to advertise on websites without explicitly stating that "the drugs are not approved, the acquisition route, information on whether there are other domestically approved drugs with the same ingredients and performance, information on the safety of the products in other countries, and they are ineligible for the relief system for sufferers from adverse drug reactions," in Japan's medical advertising guidelines [28]. All the websites were assessed based on whether these 5 items were stated.

Japan's medical advertising guidelines also prohibit exaggerated advertisements [28]. The guidelines [28] define exaggerated advertisements as advertisements that are not necessarily false but that misleads citizens by misrepresenting facts such as about the size of the facility, staffing, or the contents of the medical care provided. Moreover, "misleading citizens" is defined as "it is sufficient to say, as a matter of common sense, that there is a difference between the 'impression' or 'expectation' that citizens perceive from the advertisements and the actual contents, and it is not necessary to prove that the advertisements mislead or to show that the advertisements have actually misled citizens" [27]. Therefore, since taking GLP-1RAs is considered an auxiliary therapy for exercise and dietary therapy in the treatment of obesity, statements such as "you can lose weight with GLP-1RAs alone," "you can lose weight even if you cannot exercise," or "you do not need dietary therapy" are considered exaggerated advertisements. The coders carefully read the guidelines and assessed each statement by using a binary (0/1) scale to indicate the presence or absence of exaggerated advertisements. RO then analyzed the contents of the exaggerated advertisements and categorized them into themes according to the specific examples in the guidelines [28].

Quality Assessment Using the DISCERN Instrument

The DISCERN instrument in the Japanese version [33] was used for the quality assessment of information. The DISCERN

instrument consists of 16 items, each of which is assessed on a 5-point Likert scale (1=not at all, 5=completely). Section 1 (questions 1-8) addresses website reliability, section 2 (questions 9-15) addresses the quality of information about treatment options, and section 3 (question 16) presents the overall assessment. The total score ranged from 16 to 80, with higher scores indicating higher quality information. Several previous studies ranked the quality of websites based on their total score as follows: excellent=80-63, good=62-51, fair=50-39, poor=38-27, very poor=26-16 [35,46].

Statistical Analysis

Data are presented as means for continuous variables and counts with frequency percentages for categorical variables. Cohen κ statistic was used to measure the intercoder reliability for compliance with medical advertising guidelines and DISCERN scores between the first and second coders. Linear regression models were used to investigate the association between the specialties of the clinic operating the website and the total DISCERN score. *P* values were 2-sided, and $P<.05$ was considered statistically significant. All analyses were conducted using the R version 4.3.1 (June 16, 2023; R Foundation for Statistical Computing).

Ethical Considerations

This study did not involve human or animal participants, and all data were collected from publicly available web-based sources. This study was granted an exemption from ethics approval by the ethics review committee of the Graduate School of Medicine, University of Tokyo.

Results

Characteristics of the Included Websites

In total, 240 websites were identified. After the removal of intrasearch term duplicates and those that met the exclusion criteria, 87 websites were used for the analysis (Figure 1). A total of 67 clinics operated the 87 websites. This was because 8 clinics had multiple different GLP-1RA diet-related articles. Table 1 shows that the most common departments, as far as could be determined from the websites, were dermatology (22/120, 18.3%), cosmetic dermatology (21/120, 17.5%), and internal medicine (19/120, 15.8%); 17 clinics were unaware of their departments. The percentages of advertisements with Ozempic, Rybelsus, Mounjaro, Victoza, and Saxenda were 18.5% (30/162), 29.1% (47/162), 12.3% (20/162),14.2% (23/162), and 25.9% (42/162), respectively. Approximately 18% (16/87) of the websites mentioned selling sodium-glucose cotransporter 2 inhibitor along with GLP-1RAs, and 12% (10/87) mentioned selling metformin.

Figure 1. Search process flowchart. GLP-1: glucagon-like peptide-1; GLP-1RA: glucagon-like peptide-1 receptor agonist.

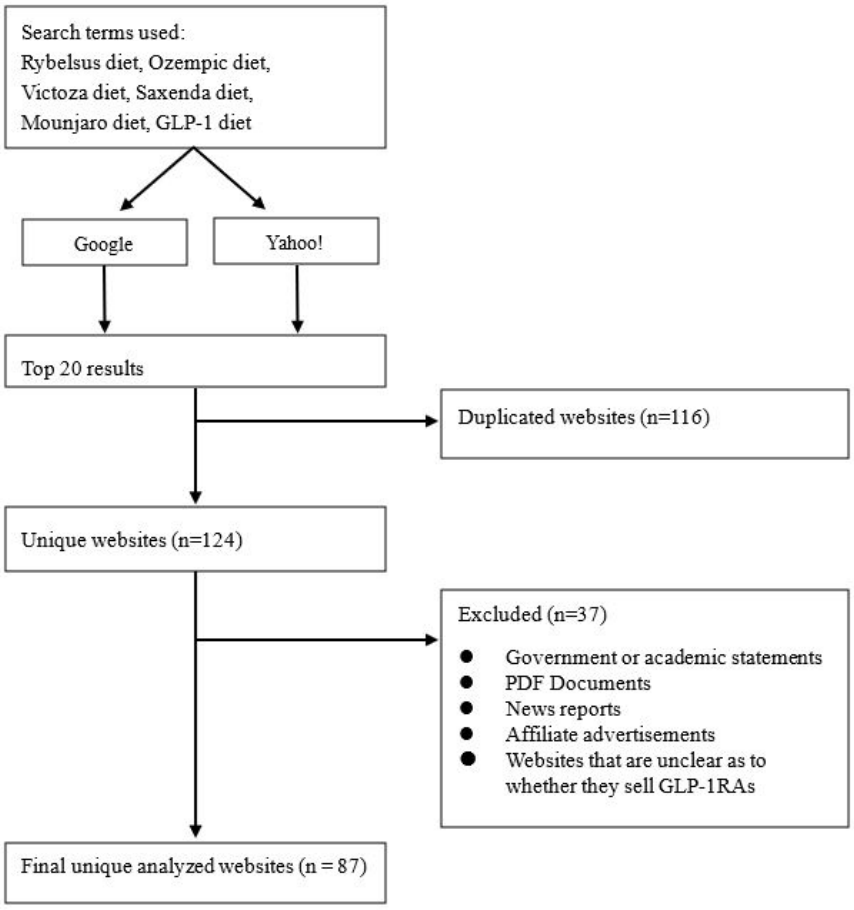


Table 1. Characteristics of the websites regarding prescription drug misuse of glucagon-like peptide-1 receptor agonists by off-label use for weight loss.

	Values, n (%)
Specialty[a] (n=120[b])	
Dermatology	22 (18.3)
Cosmetic dermatology	21 (17.5)
Internal medicine	19 (15.8)
Cosmetic surgery	6 (5)
Urology	6 (5)
Allergology	5 (4.2)
Other	24 (20)
Unknown	17 (14.2)
Types of advertised drugs[c] (n=162[d])	
Ozempic	30 (18.5)
Rybelsus	47 (29)
Mounjaro	20 (12.3)
Victoza	23 (14.2)
Saxenda	42 (25.9)

[a]Medical departments represented by medical institutions.

[b]Total number of mentions of specialty.

[c]Types of glucagon-like peptide-1 receptor agonists that medical institutions advertised for sale.

[d]Total number of mentions of drugs.

Compliance With Medical Advertising Guidelines

The intercoder reliability was acceptable (average Cohen κ=0.667). Table 2 shows the compliance with medical advertising guidelines. Only one of the 87 websites stated the 5 required items. The items with the highest percentage of listing were "acquisition route" and "information on the safety of the drugs in other countries," both at 79% (69/87). The item with the lowest percentage of listing was "ineligible for the relief system for sufferers from adverse drug reactions" at 9% (8/87).

Table 2. Compliance with advertising guidelines of the websites regarding prescription drug misuse of glucagon-like peptide-1 receptor agonists by off-label use for weight loss (N=87).

	Values, n (%)
Content that should be included[a]	
The drug is unapproved	43 (49)
Acquisition route	69 (79)
Information on whether there are other domestically approved drugs with the same ingredients and performance	24 (28)
Information on the safety of the drug in other countries	69 (79)
Ineligible for relief system for sufferers from adverse drug reactions	8 (9)
Exaggerated advertisements[b]	72 (83)

[a]Explicit statements required under Japan's medical advertising guidelines for noninsured medical treatments involving off-label drug use.

[b]Advertisements containing expressions that could mislead citizens, as determined by coders under Japan's medical advertising guidelines.

Table 2 shows that 83% (72/87) of the websites listed exaggerated advertisements. Table 3 shows the content of the exaggerated advertisements: of the 87 websites, 69% (60/87) conveyed messages indicating no exercise and dietary therapy were required, 24% (21/87) conveyed messages indicating that using GLP-1RAs is a natural and healthy method, and 31% (27/87) provided the website author's subjective expression of the risks of using GLP-1RAs.

Table 3. Content of exaggerated advertising regarding prescription drug misuse of glucagon-like peptide-1 receptor agonists by off-label use for weight loss (N=87)[a].

Themes	Illustrative quotes	Values, n (%)
Induction by nonscientific information (no diet or exercise therapy required)	*…Exercise is also unnecessary. Exercise is less effective in losing weight.* *…Medical dieting is a treatment that suppresses appetite and promotes weight loss by taking a medicine called GLP-1 receptor agonist, which is used in the treatment of diabetes, without diet restrictions or hard exercise regimen.*	60 (69)
Induction by nonscientific information (natural and healthy methods)	*…By managing portion sizes, you can reduce calorie intake and engage in a natural and controlled dietary regimen.* *…To achieve weight loss in a healthy manner.* *…It also has fat-burning and pancreas-protecting effects, making it a healthy way to lose weight.*	21 (24)
Subjective expressions of risks	*…Since the delayed gastric emptying effect is not long-lasting, there is little need to consider it dangerous.* *…I, as the director, have also used it and had no side effects other than slight heartburn, which may have caused mild anorexia and slight weight loss.*	27 (31)
Statements that you can always get the drug only through web-based medical consultation	*…Everything from making an appointment and consultation to prescribing a drug is done online, so there is no need to go out.* *…You can see us from anywhere in the country via our web-based clinic. [...] There is no consultation fee; GLP-1 injections will be mailed to you.*	20 (23)

[a]Advertisements containing expressions that could mislead citizens, as determined by coders under Japan's medical advertising guidelines.

Quality Assessment Using DISCERN Instrument

Intercoder reliability was acceptable (average Cohen κ=0.663). Figure 2 shows the DISCERN scores for all 87 websites across each item. Table 4 presents the mean score for each of the 16 DISCERN instrument criteria and the total DISCERN score. The mean total DISCERN score for all 87 websites was 32.6 (SD 5.5). Only 1 website scored more than 51 points, which is good; 9 websites scored between 39 and 50 points, which is fair; 72 websites were poor; and 5 websites were very poor. The items that scored a mean of at least 3 points were criterion 9 (does it describe how each treatment works?) with a mean score of 3.96 (SD 1.10), criterion 2 (does it achieve its aims?) with a mean score of 3.18 (SD 0.70), criterion 1 (are the aims clear?) with a mean score of 3.10 (SD 0.57), and criterion 10 (does it describe the benefits of each treatment?) with a mean score of 3.01 (SD 0.42). The items that scored a mean of less than 2 points were criterion 15 (does it provide support for shared decision-making?), criterion 12 (does it describe what would happen if no treatment was used?), criterion 7 (does it provide details of additional sources of support and information?), criterion 8 (does it refer to the areas of uncertainty?), criterion 14 (is it clear that there may be more than one possible treatment choice?), criterion 6 (is this balanced or unbiased?), criterion 4 (is it clear what sources of information were used to compile the publication?), and criterion 5 (is it clear when the information used or reported in the publication was produced?). The mean score for criteria 16, which assessed the overall quality, was 1.49 (SD 0.73). No statistically significant association was found between the specialties and the total DISCERN score (all $P>.05$).

Figure 2. DISCERN scores for each item: #1 to #16 represent the criteria numbers of the DISCERN instrument. No: identification numbers for individual websites for analysis.

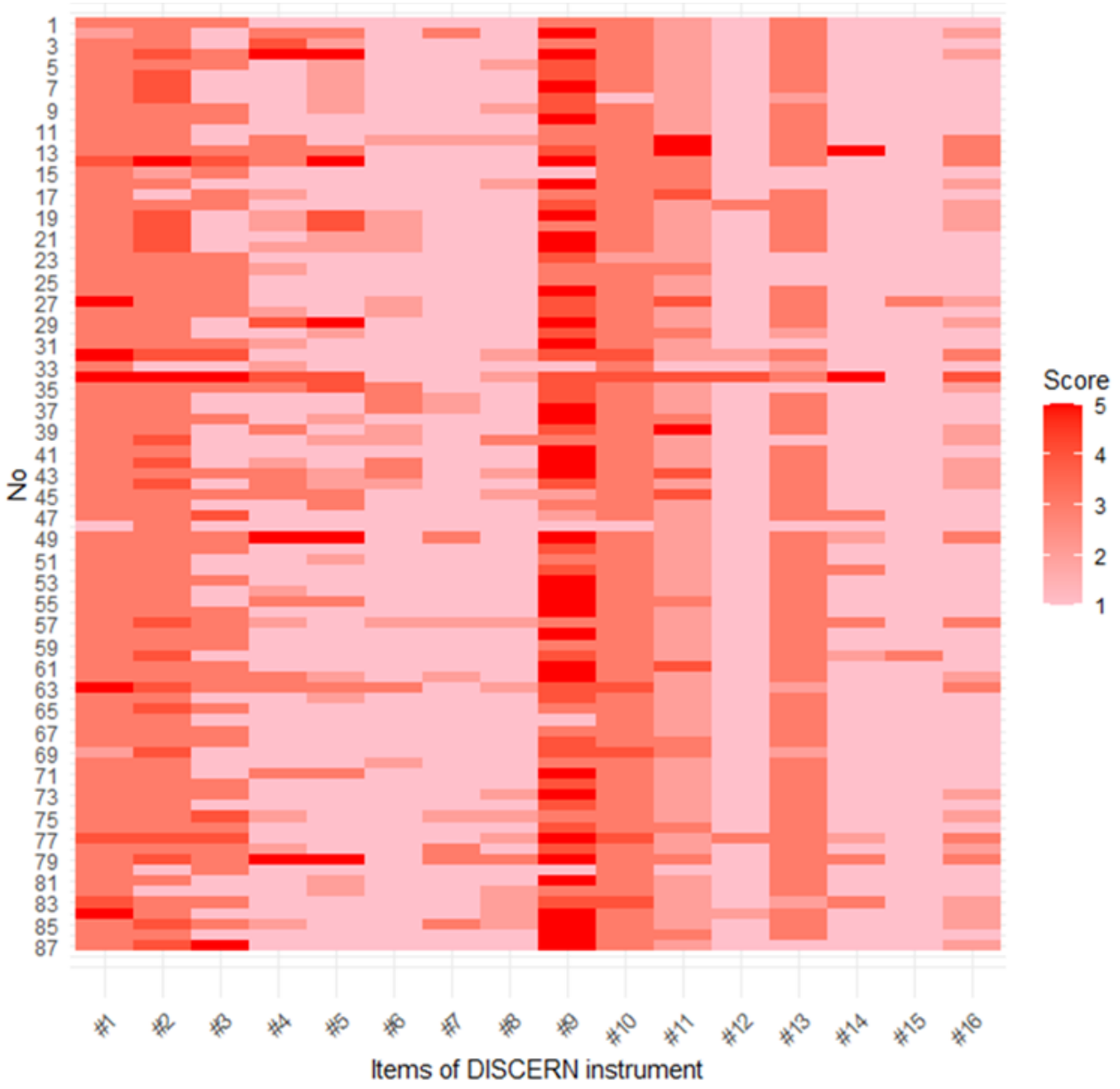


Table 4. Mean score for each of the 16 DISCERN instrument criteria[a].

Criterion number	Criteria	Mean score (SD)
1	Provides clear aims	3.10 (0.57)
2	Achieves its aims	3.18 (0.71)
3	Provides relevant information	2.21 (1.15)
4	Provides sources of information	1.70 (1.07)
5	Provides information production date	1.73 (1.16)
6	Balanced and unbiased	1.28 (0.58)
7	Provides additional sources of information	1.18 (0.52)
8	Refers to areas of uncertainty	1.24 (0.48)
9	Describes how each treatment works	3.97 (1.10)
10	Describes benefits of each treatment	3.01 (0.42)
11	Describes risks of each treatment	2.36 (0.79)
12	Describes what would happen if no treatment is used	1.10 (0.46)
13	Describes how treatment affects quality of life	2.70 (0.67)
14	Makes clear that there may be more than one possible treatment choice	1.24 (0.76)
15	Supports shared decision-making	1.05 (0.30)
16	Overall quality	1.49 (0.73)

[a]Each DISCERN item is assessed on a 5-point Likert scale (1=not at all, 5=completely). Total score = 32.55 (SD 5.53).

Discussion

Principal Findings

This study assesses compliance with Japanese medical advertising guidelines, exaggerated advertisement classification, and DISCERN scores of Japanese websites of medical institutions that prescribe GLP-1RAs off-label for weight loss, which Japanese consumers may refer to when considering weight loss with GLP-1RAs.

Many websites potentially violated Japan's medical advertising guidelines, and very few websites listed all the items required for advertising GLP-1RAs off-label use. In particular, only 9% (8/87) of the websites stated that using GLP-1RAs off-label was not covered by the Adverse Reactions Relief System for Drugs. In recent years, problems related to cosmetic therapy for weight loss have increased in Japan [47], so it is problematic that consumers have limited exposure to information indicating that the off-label use of GLP-1RAs is not covered by the Adverse Reactions Relief System for Drugs. Furthermore, concerns have been raised that GLP-1RAs may be obtained without a prescription and are at risk of misuse and abuse in inappropriate quantities [8]. However, research on who is actually misusing and what kinds of adverse effects occur because of misuse remains scarce. Generally, it has been reported that distinguishing between therapeutic errors, unintentional misuse, and intentional abuse of drugs is challenging, and the content of the reports tends to vary [5]. As mentioned above, in Japan, because it is not covered by the Adverse Reactions Relief System for Drugs, it may be difficult for consumers who have experienced adverse events due to the misuse of GLP-1RAs to claim compensation.

Approximately 83% (72/87) of the websites included exaggerated advertisements, and much of their content may mislead consumers into assuming that one can lose weight by using GLP-1RAs without exercise or dietary therapy. However, since participants in clinical trials evaluating the efficacy of obesity treatment drugs received medication therapy as an auxiliary to lifestyle interventions, the message that the use of GLP-1RAs for weight loss does not require exercise or dietary therapy is unscientific [48]. Additionally, according to another previous study, such messages may evoke a "compensatory health belief" in consumers that "if an alternative behavior is adopted, it is not necessary to engage in healthy behaviors or refrain from unhealthy ones." [49,50]

In this study, the mean of the total DISCERN scores of the analyzed websites was 32.56, which qualified as poor. Individually, most of the websites qualified as poor, only 1 website qualified as good, 9 websites qualified as fair, 72 websites qualified as poor, and 5 websites qualified as very poor. The mean score of criterion 16, the overall quality rating, was also low (score=1.49). These results were extremely low compared to those of previous studies that used the DISCERN instrument to assess the quality of health information [34-36,38]. For each question, criterion 15 (shared decision-making), criterion 12 (if no treatment is used), criterion 7 (additional information sources), criterion 8 (uncertainty), criterion 14 (possible treatment choice), and criterion 6 (balanced and unbiased) scored low, and criterion 9 (how each treatment works), criterion 2 (achieves its aims), and criterion 1 (provides clear aim) scored high. This trend was consistent with that of previous studies [34-36]. Unlike previous studies that compared DISCERN scores by the type of website operator (academic, government, or commercial agencies), only private clinic

websites that prescribe GLP-1RAs for off-label use were assessed in this study. Therefore, the DISCERN score was significantly lower in this study, possibly because the purpose of the websites was to sell GLP-1RAs and not to provide accurate information about obesity or obesity treatment to the public. In other words, these websites were advertisements designed to appeal to consumers about the effectiveness of GLP-1RAs for weight loss and not to draw their attention to other treatment options or shared decision-making with other health care providers or family members. Nevertheless, a previous study on consumer judgments of the reliability of web-based health information showed that website owners' authority had a positive impact on reliability and credibility [51]. Accordingly, if medical institutions representing specialties such as internal medicine were the owners of these websites, many consumers might trust the content of the websites. In summary, the problems with the information on medical institution websites identified by this study are not only unscientific and in violation of medical advertising guidelines but are also problematic from the perspective of public health and health promotion.

In this study, we observed that there were few public institution websites providing information on obesity treatment for consumers. To the best of our knowledge, no study has comprehensively compared the overall quality of web-based medical information between Japan and other countries. However, a previous study comparing palliative care websites in the United States and Japan reported that Japanese websites were of lower quality than the American ones [52]. A study focusing on the quality of web-based medical information within Japan has generally reported low quality [53]. This issue may be influenced by differences in the medical advertising regulations between Japan and other countries. Future studies should aim to provide a more comprehensive and international assessment of the quality of medical information.

Practical Implications

The results of this study have several implications. Regarding future studies, researchers should assess the reliability of media content, including medical institutions' websites that prescribe drugs for off-label use to contribute to better control of prescription drug misuse in the future. Many studies have investigated the consumer risk factors and preventive interventions for drug misuse. For example, a study investigating the misuse of over-the-counter drugs showed that health literacy, low education level, and misunderstanding of over-the-counter medicine instructions were correlated with the incidence of misuse [5]. Another study showed that the misuse of prescription stimulants for weight loss was positively associated with being female, lacking insurance, and having a history of lifetime misuse but not past-year misuse of prescription sedatives or tranquilizers [54]. Limited health literacy has been reported to be correlated with overweight/obesity in the first place [55], greater difficulty in the medication compliance, and higher mortality [56]. Thus, efforts to enhance consumer health literacy may be equally important as well as the content reliability of medical institutions' websites. In particular, studies on the risk of GLP-1RA misuse or off-label use are scarce and should be investigated.

Regarding future practices, this study suggests that governments should continuously monitor and provide corrective actions on websites, including unscientific content, to promote drug sales. A previous study in Japan showed that the number and volume of newspaper advertisements for dietary supplements that might mislead consumers decreased following the collective action of Japan's Consumer Affairs Agency [57]. Therefore, government monitoring and regulations are also important to address misleading information provided by medical institutions concerning prescription drug misuse. Furthermore, there are no uniform international guidelines for medical advertising. The permissibility of over-the-counter drug advertisements and advertisements for medical services differs from country to country. However, with the recent proliferation of social networking service platforms, misinformation or information with commercial intent about medicine can spread around the world in a short period of time. Therefore, it is also important to establish international medical advertising guidelines. Second, medical institutions and health care providers should recognize the current problem of GLP-1RA misuse and provide accurate information on the internet and through in-person explanations to individuals who are interested in weight loss. Additionally, medical institutions, academic organizations, and ministries should improve their websites with a focus on search engine optimization to ensure that high-quality medical information appears at the top of search engine results [58]. Internet-based companies, including Google, should be encouraged to prioritize the display of high-quality medical information. Third, medical institutions, health care providers, academic institutions, and ministries should inform citizens that current information on GLP-1RAs on websites is unreliable and that there is no miracle drug that is risk-free and absolutely effective. If individuals without obesity wish to lose weight, they should consult a trusted specialist such as a primary care physician to understand the risks and proceed with the therapy best suited for them.

Limitations and Strengths

This study has several limitations. First, this study only analyzed Japanese-language websites. The approval status of GLP-1RAs and medical advertising regulations differ from country to country. Thus, the generalizability of the results of this study to other countries is limited. Second, this study assesses the search results for web browsers and websites when viewed with web browsers. The results and website display contents may differ when using a smartphone. The results may have been affected by websites and search algorithms that were optimized for smartphones. Third, this study only used the DISCERN instrument for quality assessment. Therefore, this study was unable to discuss the reliability items other than those evaluated by DISCERN. There are multiple other quality assessment indicators such as the Health on the Net Foundation Code of Conduct certification and the Journal of the American Medical Association (JAMA) benchmark criteria. However, the Health on the Net Foundation Code of Conduct is not well known in Japan, and few websites have obtained it; therefore, it was not used in this study. The JAMA benchmark has only 4 criteria. Since most of the medical institutions of this study were private clinics, it was assumed that the 2 criteria of JAMA benchmark, authorship and attribution, would not differ. Additionally,

previous studies that evaluated web-based health information reported a positive correlation between JAMA benchmark scores and DISCERN scores [59,60]. Therefore, only DISCERN was used in this study. Additionally, the DISCERN instrument was developed to enable patients to assess the quality of written information related to treatment choices and to promote the creation of high-quality health information for documents intended for patients. Thus, assessing quality in terms of medical accuracy is not the primary focus of the DISCERN instrument. However, in this study, 2 internal medicine physician coders assessed the medical accuracy when evaluating questions 6, 9, and 12 of the DISCERN instrument, thereby reflecting medical accuracy in the DISCERN score. Fourth, this study relies on a single reviewer to screen eligible websites, which raises the possibility that some potentially eligible websites may have been excluded from the analyzed dataset. Fifth, the intercoder reliability between the 2 coders was ensured to be acceptable; however, coder bias may not have been completely eliminated. Finally, the impact of the websites analyzed in this study on viewers' cognition and behavior was outside the scope of this study; future studies should examine this aspect.

Despite these limitations, to the best of our knowledge, this is the first study on the quality of information on the websites of medical institutions that prescribe GLP-1RAs for weight loss in Japan. This study provides important implications for future studies and practice on better communication to decrease prescription drug misuse of GLP-1RAs, as discussed above. The quality of the websites of medical institutions that prescribe GLP-1RAs for weight loss was very low, and many of them did not comply with the medical advertising guidelines in Japan. This study shows that many websites, including exaggerated advertisements, may promote the misconception that exercise and dietary therapy are unnecessary. This study also shows that many websites provide biased information that may lead consumers to use GLP-1RAs without presenting alternative weight loss methods other than GLP-1RAs or the option of shared decision-making with health care providers or family members. Consumers and citizens who read such information may misuse GLP-1RAs because of insufficient understanding of their side effect risks and proper usage. Public institutions must monitor and regulate advertising content that violates guidelines and misleads individuals. Health care providers need to inform consumers about the risks and proper usage of GLP-1RAs and provide accurate information related to weight loss. Beneficial medications developed for people's health should not pose health risks because of the inappropriate information provided by health care providers.

Acknowledgments

This work was supported by the Japan Society for the Promotion of Science KAKENHI (23K09595).

Data Availability

The data presented in this study are available upon reasonable request to the corresponding author.

Authors' Contributions

RO and TO conceptualized this study and performed the methodology. RO and EF performed the formal analysis. RO wrote the original draft. HO acquired the funds. TK supervised this study. All authors contributed to reviewing and editing the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
DISCERN instrument.
[DOC File , 29 KB - formative_v9i1e68792_app1.doc]

References

1. Casati A, Sedefov R, Pfeiffer-Gerschel T. Misuse of medicines in the European Union: a systematic review of the literature. Eur Addict Res 2012;18(5):228-245 [FREE Full text] [doi: 10.1159/000337028] [Medline: 22572594]
2. Schifano F, Chiappini S, Miuli A, Mosca A, Santovito MC, Corkery JM, et al. Focus on over-the-counter drugs? misuse: A systematic review on antihistamines, cough medicines, and decongestants. Front Psychiatry 2021;12:657397 [FREE Full text] [doi: 10.3389/fpsyt.2021.657397] [Medline: 34025478]
3. Martins SS, Sampson L, Cerdá M, Galea S. Worldwide prevalence and trends in unintentional drug overdose: A systematic review of the literature. Am J Public Health 2015 Nov;105(11):e29-e49. [doi: 10.2105/AJPH.2015.302843] [Medline: 26451760]
4. Smith SM, Dart RC, Katz NP, Paillard F, Adams EH, Comer SD, et al. Classification and definition of misuse, abuse, and related events in clinical trials: ACTTION systematic review and recommendations. Pain 2013 Nov;154(11):2287-2296 [FREE Full text] [doi: 10.1016/j.pain.2013.05.053] [Medline: 23792283]

5. Algarni M, Hadi MA, Yahyouche A, Mahmood S, Jalal Z. A mixed-methods systematic review of the prevalence, reasons, associated harms and risk-reduction interventions of over-the-counter (OTC) medicines misuse, abuse and dependence in adults. J Pharm Policy Pract 2021 Sep 13;14(1):76 [FREE Full text] [doi: 10.1186/s40545-021-00350-7] [Medline: 34517925]
6. Hashimoto T, Kaneda Y, Ozaki A, Hori A, Tsuchiya T. Eleven-year trend of drug and chemical substance overdose at a local emergency hospital in Japan. Cureus 2022 Dec;14(12):e32475 [FREE Full text] [doi: 10.7759/cureus.32475] [Medline: 36644086]
7. Matsumoto T. Distributed research report: survey of drug-related psychiatric disorders in psychiatric hospitals in Japan. Department of Drug Dependence Research.: National Institute of Mental Health; 2020. URL: https://www.ncnp.go.jp/nimh/yakubutsu/report/pdf/J_NMHS_2020.pdf [accessed 2024-12-22]
8. Chiappini S, Vickers-Smith R, Harris D, Papanti Pelletier GD, Corkery JM, Guirguis A, et al. Is there a risk for semaglutide misuse? focus on the Food and Drug Administration's FDA Adverse Events Reporting System (FAERS) pharmacovigilance dataset. Pharmaceuticals (Basel) 2023 Jul 11;16(7):994 [FREE Full text] [doi: 10.3390/ph16070994] [Medline: 37513906]
9. Butuca A, Dobrea CM, Arseniu AM, Frum A, Chis AA, Rus LL, et al. An assessment of semaglutide safety based on real world data: from popularity to spontaneous reporting in Eudravigilance database. Biomedicines 2024 May 18;12(5):1124 [FREE Full text] [doi: 10.3390/biomedicines12051124] [Medline: 38791086]
10. Mahase E. Semaglutide: FDA publishes warning amid reports of patients overdosing. BMJ 2024 Jul 30;386:q1696. [doi: 10.1136/bmj.q1696] [Medline: 39079728]
11. Han SH, Safeek R, Ockerman K, Trieu N, Mars P, Klenke A, et al. Public interest in the off-label use of glucagon-like peptide 1 agonists (Ozempic) for cosmetic weight loss: a Google trends analysis. Aesthet Surg J 2023 Dec 14;44(1):60-67. [doi: 10.1093/asj/sjad211] [Medline: 37402640]
12. FDA approves new drug treatment for chronic weight management, first since 2014. USFDA. URL: https://www.fda.gov/news-events/press-announcements/fda-approves-new-drug-treatment-chronic-weight-management-first-2014 [accessed 2024-09-08]
13. Suran M. As Ozempic's popularity soars, here's what to know about semaglutide and weight loss. JAMA 2023 May 16;329(19):1627-1629. [doi: 10.1001/jama.2023.2438] [Medline: 37099334]
14. Goodman B. Poison centers see nearly 1,500% increase in calls related to injected weight-loss drugs as people accidentally overdose. CNN Health. URL: https://www.cnn.com/2023/12/13/health/semaglutide-overdoses-wellness [accessed 2024-10-21]
15. Brafman N. Ozempic: French authorities issue alert for anti-diabetic drug misused for weight loss. Le Monde 2023 Mar 2. URL: https://www.lemonde.fr/en/health/article/2023/03/02/ozempic-french-authorities-issue-alert-for-anti-diabetic-drug-misused-for-weight-loss_6017913_14.html [accessed 2024-10-21]
16. Mahase E. GLP-1 shortages will not resolve this year, EMA warns, amid concern over off-label use. BMJ 2024 Jun 28;385:q1448. [doi: 10.1136/bmj.q1448] [Medline: 38942431]
17. Hathaway JT, Shah MP, Hathaway DB, Zekavat SM, Krasniqi D, Gittinger JW, et al. Risk of nonarteritic anterior ischemic optic neuropathy in patients prescribed semaglutide. JAMA Ophthalmol 2024 Aug 01;142(8):732-739. [doi: 10.1001/jamaophthalmol.2024.2296] [Medline: 38958939]
18. Sodhi M, Rezaeianzadeh R, Kezouh A, Etminan M. Risk of gastrointestinal adverse events associated with glucagon-like peptide-1 receptor agonists for weight loss. JAMA 2023 Nov 14;330(18):1795-1797 [FREE Full text] [doi: 10.1001/jama.2023.19574] [Medline: 37796527]
19. Recommendations of the executive steering group on shortages and safety of medicinal products on shortage of glucagon-like peptide-1 (GLP-1) receptor agonists. European Medicines Agency. URL: https://www.quotidianosanita.it/allegati/allegato1719409581.pdf [accessed 2024-09-08]
20. Notes on the establishment of guidelines for the promotion of appropriate use of semaglutide (genetic recombination) preparation for the indication of obesity document in Japanese. Ministry of Health, Labour and Welfare, Japan. URL: https://www.pmda.go.jp/files/000265450.pdf [accessed 2024-09-08]
21. Hakariya H, Ohnishi M, Tanimoto T. Japan initiates market authorization of weight-loss drug semaglutide under universal health coverage, but with stringent prescription restrictions. Diabetes Obes Metab 2024 Aug;26(8):3006-3008. [doi: 10.1111/dom.15638] [Medline: 38747140]
22. Request for cooperation due to tight inventory of GLP-1 receptor agonist (Part 2) document in Japanese. Ministry of Health, Labour, and Welfare, Japan. URL: https://www.mhlw.go.jp/content/001165743.pdf [accessed 2024-11-09]
23. Position on off-label use of GLP-1 receptor agonists and GIP/GLP-1 receptor agonists document in Japanese. The Japan Diabetes Society. URL: https://www.jds.or.jp/uploads/files/document/info/jds_statement_GLP-1.pdf [accessed 2024-10-28]
24. Statement on the safe and appropriate use of obesity drugs document in Japanese. Japan Society for the Study of Obesity. URL: http://www.jasso.or.jp/data/Introduction/pdf/academic-information_statement_20231127.pdf [accessed 2024-09-08]
25. Telemedicine trouble for slimming purposes: Subscription trouble? where diabetes drugs are prescribed off-label for several months for the purpose of weight loss document in Japanese. National Consumer Affairs Center of Japan. URL: https://www.kokusen.go.jp/news/data/n-20231220_1.html [accessed 2024-09-08]

26. Before you get a cosmetic procedure, recheck to make sure you've been properly briefed and understand it again document in Japanese. Consumer Affairs Agency, Government of Japan. URL: https://www.caa.go.jp/policies/policy/consumer_policy/information/information_002/assets/information_002_240202_02.pdf [accessed 2024-09-08]
27. Medical Care Act (Japanese Law Translation System). Ministry of Justice, Japan. 1948. URL: https://www.japaneselawtranslation.go.jp/ja/laws/view/4006 [accessed 2024-09-08]
28. Regulations on advertisements of hospitals under the Medical Care Act (in Japanese). Ministry of Health, Labour and Welfare, Japan. URL: https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kenkou_iryou/iryou/kokokukisei/index.html [accessed 2024-09-08]
29. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health 1999 Feb;53(2):105-111 [FREE Full text] [doi: 10.1136/jech.53.2.105] [Medline: 10396471]
30. Shan Y, Xing Z, Dong Z, Ji M, Wang D, Cao X. Translating and adapting the DISCERN instrument into a simplified Chinese version and validating its reliability: Development and usability study. J Med Internet Res 2023 Feb 02;25:e40733 [FREE Full text] [doi: 10.2196/40733] [Medline: 36729573]
31. Montoya A, Llopis N, Gilaberte I. Validation of the translation of an instrument to measure reliability of written information on treatment choices: a study on attention deficit/hyperactivity disorder (ADHD). Educ Health (Abingdon) 2011 Dec;24(3):577. [Medline: 22267353]
32. Logullo P, Torloni MR, de O C Latorraca C, Riera R. The Brazilian Portuguese version of the DISCERN instrument: Translation procedures and psychometric properties. Value Health Reg Issues 2019 Dec;20:172-179 [FREE Full text] [doi: 10.1016/j.vhri.2019.09.001] [Medline: 31622803]
33. Batchelor JM, Ohya Y. Use of the DISCERN instrument by patients and health professionals to assess information resources on treatments for asthma and atopic dermatitis. Allergol Int 2009 Mar;58(1):141-145 [FREE Full text] [doi: 10.2332/allergolint.08-SC-0022] [Medline: 19153541]
34. Killip SC, Kwong NKR, MacDermid JC, Fletcher AJ, Carleton NR. The quality, readability, completeness, and accuracy of PTSD websites for firefighters. Int J Environ Res Public Health 2020 Oct 19;17(20):7629 [FREE Full text] [doi: 10.3390/ijerph17207629] [Medline: 33086772]
35. Li Y, Zhou X, Zhou Y, Mao F, Shen S, Lin Y, et al. Evaluation of the quality and readability of online information about breast cancer in China. Patient Educ Couns 2021 Apr;104(4):858-864 [FREE Full text] [doi: 10.1016/j.pec.2020.09.012] [Medline: 32988687]
36. Tan DJY, Ko TK, Fan KS. The readability and quality of web-based patient information on nasopharyngeal carcinoma: quantitative content analysis. JMIR Form Res 2023 Nov 27;7:e47762 [FREE Full text] [doi: 10.2196/47762] [Medline: 38010802]
37. Patel AA, Joshi C, Varghese J, Hassan AM, Janis JE, Galiano RD. Do websites serve our patients well? a comparative analysis of online information on cosmetic injectables. Plast Reconstr Surg 2022 Apr 01;149(4):655e-668e. [doi: 10.1097/PRS.0000000000008921] [Medline: 35139065]
38. Daraz L, Morrow AS, Ponce OJ, Beuschel B, Farah MH, Katabi A, et al. Can patients trust online health information? a meta-narrative systematic review addressing the quality of health information on the internet. J Gen Intern Med 2019 Sep;34(9):1884-1891 [FREE Full text] [doi: 10.1007/s11606-019-05109-0] [Medline: 31228051]
39. Weaver JB, Mays D, Weaver SS, Hopkins GL, Eroglu D, Bernhardt JM. Health information-seeking behaviors, health indicators, and health risks. Am J Public Health 2010 Aug;100(8):1520-1525. [doi: 10.2105/AJPH.2009.180521] [Medline: 20558794]
40. Telecommunications usage trends survey (in Japanese). Ministry of Internal Affairs and Communications. URL: https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r05/html/datashu.html#f00281 [accessed 2024-10-12]
41. Montemurro P, Porcnik A, Hedén P, Otte M. The influence of social media and easily accessible online information on the aesthetic plastic surgery practice: literature review and our own experience. Aesthetic Plast Surg 2015 Apr;39(2):270-277. [doi: 10.1007/s00266-015-0454-3] [Medline: 25697277]
42. Google Japan. URL: https://www.google.co.jp/ [accessed 2024-12-23]
43. Yahoo! Japan. URL: https://www.yahoo.co.jp/ [accessed 2024-12-23]
44. Search engine share in Japan. Ministry of Internal Affairs and Communications. URL: https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r05/html/datashu.html#f00228 [accessed 2024-09-08]
45. Eysenbach G, Köhler C. How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. BMJ 2002 Mar 09;324(7337):573-577 [FREE Full text] [doi: 10.1136/bmj.324.7337.573] [Medline: 11884321]
46. Ved R, Cobbold N, Igbagiri K, Willis M, Leach P, Zaben M. Online patient information on Vagus Nerve Stimulation: How reliable is it for facilitating shared decision making? Seizure 2017 Aug;50:125-129 [FREE Full text] [doi: 10.1016/j.seizure.2017.06.009] [Medline: 28651125]
47. Trouble with online medical treatment for slimming purposes, etc document in Japanese. National Consumer Affairs Center, Japan. URL: https://www.kokusen.go.jp/pdf/n-20231220_1.pdf [accessed 2024-10-01]

48. Jastreboff AM, Aronne LJ, Ahmad NN, Wharton S, Connery L, Alves B, SURMOUNT-1 Investigators. Tirzepatide once weekly for the treatment of obesity. N Engl J Med 2022 Jul 21;387(3):205-216. [doi: 10.1056/NEJMoa2206038] [Medline: 35658024]
49. Hartmann C, Keller C, Siegrist M. Compensatory beliefs, nutrition knowledge and eating styles of users and non-users of meal replacement products. Appetite 2016 Oct 01;105:775-781. [doi: 10.1016/j.appet.2016.07.013] [Medline: 27421243]
50. Iye R, Okuhara T, Okada H, Goto E, Furukawa E, Kiuchi T. The effect of exposure to "exemption" video advertisements for functional foods: a randomized control study in Japan. Healthcare (Basel) 2022 Feb 11;10(2):345 [FREE Full text] [doi: 10.3390/healthcare10020345] [Medline: 35206959]
51. Zhang Y, Kim Y. Consumers' evaluation of web-based health information quality: Meta-analysis. J Med Internet Res 2022 Apr 28;24(4):e36463 [FREE Full text] [doi: 10.2196/36463] [Medline: 35482390]
52. Tanabe K, Fujiwara K, Ogura H, Yasuda H, Goto N, Ohtsu F. Quality of web information about palliative care on websites from the United States and Japan: comparative evaluation study. Interact J Med Res 2018 Apr 03;7(1):e7 [FREE Full text] [doi: 10.2196/ijmr.9574] [Medline: 29615388]
53. Yamamura A, Watanabe S, Yamaguchi S, Iwata K, Kimura S, Mikami Y, et al. Readability and quality of online patient resources regarding knee osteoarthritis and lumbar spinal stenosis in Japan. J Orthop Sci 2024 Sep;29(5):1313-1318. [doi: 10.1016/j.jos.2023.08.003] [Medline: 37599135]
54. Compton WM, Han B, Blanco C, Johnson K, Jones CM. Prevalence and correlates of prescription stimulant use, misuse, use disorders, and motivations for misuse among adults in the United States. Am J Psychiatry 2018 Aug 01;175(8):741-755 [FREE Full text] [doi: 10.1176/appi.ajp.2018.17091048] [Medline: 29656665]
55. Michou M, Panagiotakos DB, Costarelli V. Low health literacy and excess body weight: a systematic review. Cent Eur J Public Health 2018 Sep;26(3):234-241 [FREE Full text] [doi: 10.21101/cejph.a5172] [Medline: 30419628]
56. Fan Z, Yang Y, Zhang F. Association between health literacy and mortality: a systematic review and meta-analysis. Arch Public Health 2021 Jul 01;79(1):119 [FREE Full text] [doi: 10.1186/s13690-021-00648-7] [Medline: 34210353]
57. Okuhara T, Yokota R, Shirabe R, Iye R, Okada H, Kiuchi T, et al. Japanese newspaper advertisements for dietary supplements before and after COVID-19: a content analysis. BMJ Open 2021 Nov 23;11(11):e050898 [FREE Full text] [doi: 10.1136/bmjopen-2021-050898] [Medline: 34815281]
58. Modave F, Shokar NK, Peñaranda E, Nguyen N. Analysis of the accuracy of weight loss information search engine results on the internet. Am J Public Health 2014 Oct;104(10):1971-1978. [doi: 10.2105/AJPH.2014.302070] [Medline: 25122030]
59. Foster B, Grampp S, Ozdag Y, Tang A, Vazquez F, Mercuri J. Assessing the quality, content, and readability of online patient resources on viscosupplementation. Journal of Orthopaedic Experience & Innovation Charter Services New York d/b/a Journal of Orthopaedic Experience and Innovation; 2024 Oct 4 2024 Oct 4;5(2):1-8. [doi: 10.60118/001c.120353]
60. Ozduran E, Büyükçoban S. Evaluating the readability, quality and reliability of online patient education materials on post-covid pain. PeerJ 2022;10:e13686 [FREE Full text] [doi: 10.7717/peerj.13686] [Medline: 35880220]

Abbreviations

FDA: Food and Drug Administration
GLP-1: glucagon-like peptide-1
GLP-1RA: glucagon-like peptide-1 receptor agonist
JAMA: Journal of the American Medical Association

Edited by A Mavragani; submitted 13.11.24; peer-reviewed by E Gray, N Misra, M Agbede; comments to author 30.11.24; revised version received 08.12.24; accepted 16.12.24; published 01.01.25.

Please cite as:
Oyama R, Okuhara T, Furukawa E, Okada H, Kiuchi T
Quality Assessment of Medical Institutions' Websites Regarding Prescription Drug Misuse of Glucagon-Like Peptide-1 Receptor Agonists by Off-Label Use for Weight Loss: Website Evaluation Study
JMIR Form Res 2025;9:e68792
URL: https://formative.jmir.org/2025/1/e68792
doi:10.2196/68792
PMID:

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org